MULLARD LIMITED
Mullard House
Torrington Place
London, W.C.1

CONTENTS

Gallium Arsenide Semiconductor Devices 118–126

Transient Performance of Heatsinks for Power Devices 127–130

Operating Principles of the Z504S Stepping Tube 131–143

VOL. 8 No. 75 MARCH 1965

Mullard Technical Communications

VOLUME 8 NUMBER 75

March 1965

Mullard Technical Communications

VOLUME 8 NUMBER 75 March 1965

Contents

Gallium Arsenide Semiconductor Devices
J. S. Lamming and C. A. P. Foxell 118

Transient Performance of Heatsinks for Power Devices
J. H. Tuley 127

Operating Principles of the Z504S Stepping Tube
C. S. Barker 131

Mullard Technical Communications contains information on the performance or application of Mullard products such as electronic valves and tubes, semiconductor devices, and electrical and magnetic components. It is issued free-of-charge to equipment designers, research organisations, reference libraries, and educational establishments. Bona fide applications to be added to the circulation list should be made under an official letter head to Central Technical Services. Overseas enquiries should be addressed to the local Mullard agent, or to Mullard Overseas Ltd. at Mullard House.

Abstracts, or references to an article, should mention both the source, *Mullard Technical Communications*, and the authors. Requests for republication or translation should be made to the Editor in writing.

The issue of the information contained in this publication does not imply any authority or licence for the utilisation of any patented feature.

'Mullard' is the trade mark of Mullard Limited and is registered in most of the principal countries of the world.

Editor: F. A. Sowan

Assistant Editors: R. Barrie E. G. Evans

Published by
Central Technical Services
Mullard Limited, Mullard House, Torrington Place, London, W.C.1
LANgham 6633

© Mullard Limited, 1965

Made and Printed in England by
WIGHTMAN AND COMPANY LIMITED, 1/3 BRIXTON ROAD, LONDON, S.W.9

Gallium Arsenide Semiconductor Devices

J. S. LAMMING and C. A. P. FOXELL
Wembley Laboratories, Associated Semiconductor Manufacturers Limited

After the successful use of the elemental semiconductors, considerable attention has been directed to developing other semiconducting materials. Of these gallium arsenide, which is one of the III-V compounds, has shown considerable promise of improving the performance of existing devices. Compared with germanium and silicon, gallium arsenide has certain attractive properties. A larger energy gap gives the possibility of high-temperature operation, higher mobility offers better high-frequency performance, and a low lifetime should result in superior switching characteristics. The first part of the article describes how far these improvements have been realised in the field of computer, varactor, mixer, tunnel and heterojunction diodes. The progress on gallium arsenide transistors and solar cells is then discussed. In addition, gallium arsenide has other interesting properties, indicating that a new class of solid-state devices based on optical mechanisms is feasible.

INTRODUCTION

Following the successful use of the elemental semiconductors germanium and silicon, considerable research has been directed to discovering other semiconducting materials with suitable properties for application in electronic devices.

Welker (Ref. 1) in 1952 showed that the Group III–Group V compounds had similar characteristics to the Group IV semiconductors, and his work started a widespread activity on these new materials. Initially, the effort was directed at using the new semiconductors as a possible method of improving the performance of existing types of solid-state device, but with an increasing knowledge of their properties a host of new devices is now envisaged. Although the compound semiconductors have interesting properties, their actual preparation in the form of single crystals of reasonable purity has met with considerable difficulties, and this more than any other factor has retarded progress. However, techniques have gradually been evolved for producing some materials reliably and in reasonable quantity. Of these compounds, gallium arsenide has emerged as a material with extremely interesting properties which not only offer a potential superiority over germanium and silicon in certain devices, but also point to the possibility of a new class of solid-state devices depending on optical mechanisms.

GALLIUM ARSENIDE

Gallium arsenide, like most of the other III–V compounds, has a zinc-blende crystal structure as shown in Fig. 1. The gallium (Group III) and arsenic (Group V) atoms occupy alternate sites in the crystal lattice, and this asymmetry is the cause of many of the material's special characteristics.

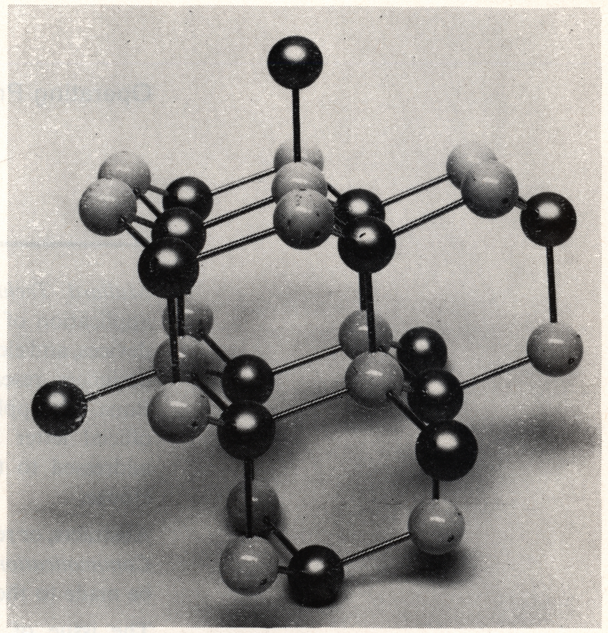


Fig. 1—Model of the crystal structure of gallium arsenide

The basic techniques evolved for producing germanium in a pure single-crystal form have been used for preparing the III–V compounds, but owing to the dissimilar characteristics of the two elements considerable practical problems are involved. In gallium arsenide the arsenic has a high vapour pressure at the temperature needed to form the compound, and the vapour pressure around the material must be adjusted to equal the dissociation pressure. High-purity gallium and arsenic are readily obtained, and large single crystals of gallium arsenide can be produced by either pulling vertically from the melt or gradient freezing by the Bridgman method. Smaller crystals can be grown from the vapour phase.

The main properties of gallium arsenide are compared with those of germanium and silicon in Table 1, and this shows that a larger energy gap, higher electron mobility, and lower lifetime are found.

TABLE 1 Comparative Properties of Gallium Arsenide

Property		Gallium arsenide	Silicon	Germanium
Melting point	(°C)	1237	1420	958
Energy gap, E_g (at 300°K)	(eV)	1·40	1·08	0·66
Dielectric constant, ε		11·1	11·8	16·0
Hole mobility, μ_p* (majority carrier)	(cm²/Vsec)	290	240	1100
Electron mobility, μ_n* (majority carrier)	(cm²/Vsec)	4500	525	2100
Avalanche breakdown voltage for an abrupt junction, V_B*	(V)	12	12	5
Lifetime, τ*	(sec)	$10^{-9} \sim 10^{-10}$	$\sim 10^{-6}$	6×10^{-5}
Maximum operating temperature	(°C)	~ 400	150	85 to 100

*corresponding to a doping level of 10^{17} cm^{-3}

The wider energy gap offers the possibility of high-temperature operation, and this in combination with a high mobility is outstanding among semiconductor materials. Direct recombination across the energy gap can take place in gallium arsenide, and although it leads to a short minority-carrier lifetime, this process has important implications for optical solid-state devices.

Another unusual feature is that a form of gallium arsenide with extremely high resistivity ($10^6 \Omega$ cm) can be produced by a compensated doping technique. This material is termed 'semi-insulating' and may well prove to be of use as a substrate for device elements in which electrical isolation is required.

There are many technological difficulties involved when solid-state devices are made from gallium arsenide. The diffusion of most of the useful impurities in gallium arsenide does not follow Fick's law, and is further complicated by the fact that the arsenic vapour pressure over the system is an independent variable which alters the diffusion coefficient. Many of the impurities form compounds or glassy layers on the surface when diffusion is attempted, leaving unsuitable surfaces for device fabrication. Alloying is also a problem. The alloying temperature must be held below about 550°C to prevent the gallium arsenide surface from dissociating. Many of the metals used have different thermal-expansion coefficients from that of gallium arsenide and form brittle eutectics which crack on cooling. Many are difficult to wet to the gallium arsenide surface, and some have inconveniently high vapour pressures. A suitable alloy which will withstand the very high operating temperatures which are possible with gallium arsenide is very difficult to find. A solution to some of these alloying problems undoubtedly lies in the use of bismuth and some of its alloys (Ref. 2).

DIODES

The high electron mobility and large energy gap of gallium arsenide offer no advantage for high-current rectifier applications, and there has been no appreciable development in this field. Owing to the wide energy gap a poor forward characteristic is obtained at low currents in comparison with silicon rectifiers, and at higher currents the lack of conductivity modulation of the base region by minority carriers increases the forward voltage drop still further. Although the large energy gap reduces the reverse leakage current and allows a higher operating temperature than with silicon, this has not yet been realised in practice owing to temperature limitations in the materials used for making alloyed contacts to the junction.

However, in the sphere of switching devices the inherently low lifetime of gallium arsenide can be turned into an advantage. When a diode is switched from the forward to the reverse direction a large recovery transient can occur, owing to the minority carriers stored in the base region diffusing back across the junction. This stored charge is in addition to the capacitive charge of the junction and can severely limit the switching performance of germanium and silicon devices. That is

$$Q_T = Q_D + Q_C \qquad \ldots(1)$$

and it can be shown that when the diffusion length of the minority carriers is less than the base width, then

$$Q_D = \tau I_F \qquad \ldots(2)$$

where

Q_T = total stored charge
Q_D = charge due to minority carriers
Q_C = capacitive charge
τ = minority-carrier lifetime
I_F = forward current.

Therefore by utilising the short lifetime of gallium

This article appeared originally in the G.E.C. Journal of Science and Technology, Vol. 31, No. 3, 1964.

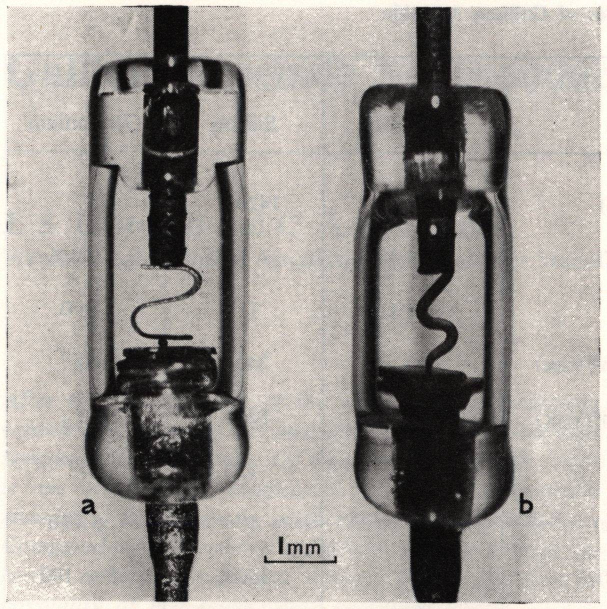


Fig. 2—(a) Alloyed-junction switching diode
(b) Point-contact switching diode

arsenide it is possible to reduce the minority-carrier storage to negligible proportions. This has been exploited in an experimental alloyed-junction switching diode of this type (Fig. 2a) which has been produced in these laboratories. The typical electrical characteristics given in Figs. 3 and 4 and Table 2, show that this diode has attractive characteristics and that the small recovery transient is entirely due to the junction capacitance.

In an effort to reduce this capacitance a smaller-area diode has been evolved as shown in Fig. 2b. By combining the low capacitance of a point-contact structure with the insignificant minority-carrier storage of gallium arsenide, a diode with a recovery time of less than 1 nanosecond has been achieved. The forward characteristic is better than that found with other types of gallium arsenide diode, the junction being formed by a metal-semiconductor contact, and it becomes comparable with those of existing germanium and silicon diodes. From the characteristics shown in Figs. 3 and 4 and Table 3, it can be seen that this diode possesses excellent static and switching characteristics for computer and high-frequency applications. This work has resulted in a diode of similar performance, available under the development type number 54CAY.

VARACTOR DIODES

As a result of the introduction of silicon variable-capacitance diodes and their use in parametric amplifiers and harmonic generators, there has been an increasing demand for varactor diodes capable of operating in the microwave region. The cut-off frequency is normally defined by

$$f_c = \frac{1}{2\pi RC} \qquad \ldots(3)$$

where

R = total series resistance (usually measured at X-band)
C = junction capacitance (usually at −6V).

It can be shown, on the basis of simple diode geometry, that in terms of the semiconductor material parameters this can be expressed as

$$f_c \propto \mu \varepsilon^{-\frac{1}{2}} \qquad \ldots(4)$$

where

μ = majority-carrier mobility
ε = dielectric constant.

As gallium arsenide has a similar dielectric constant to silicon but a higher mobility of 3500cm^2/Vsec, compared to 500cm^2/Vsec for typical varactor doping levels, it should be possible to obtain a considerable improvement in performance.

However, in addition to high cut-off frequency, a varactor diode must have certain other properties. Thus for high-power harmonic generation it is necessary that it should be able to withstand a large voltage swing, and also that in the forward conduction condition it should have a large diffusion capacitance (Ref. 3). Unfortunately, gallium arsenide possesses neither of these attributes at present, and its significance in harmonic generators is likely to be at low-power levels and at frequencies beyond the range available with silicon diodes.

For very-low-noise parametric amplifiers two important requirements are high cut-off frequency and operation at low temperature. In silicon the series resistance becomes large at low temperatures, whereas gallium arsenide diodes have been known to give excellent performance even at liquid-helium temperatures. It has been reported (Ref. 4) from the Bell Telephone Laboratories, where the Telstar project provides considerable impetus for low-noise amplifier development, that a gallium arsenide diode at liquid-helium temperature enabled an overall noise temperature of 23·5°K to be achieved for a practical amplifier; it is believed that these low-noise parametric amplifiers using gallium arsenide varactor diodes will compare favourably with masers, particularly since their performance is not critically dependent on operating temperature.

TABLE 2
Typical Characteristics of an Alloyed-junction Diode

Breakdown voltage	(100μA, 25°C)	23V
Reverse current	(−10V, 25°C)	30nA
Reverse current	(−10V, 100°C)	50nA
Forward voltage drop	(1mA, 25°C)	0·88V
Forward voltage drop	(100mA, 25°C)	1·45V
Forward voltage drop	(450mA, 25°C)	3·0V
Capacitance	(−2V)	3·8pF
Total recovered charge	(+10mA to −10V)	25pC

TABLE 3
Typical Characteristics of a Point-contact Switching Diode

Breakdown voltage	(100μA, 25°C)	16V
Reverse current	(−3V, 25°C)	10nA
Reverse current	(−3V, 100°C)	600nA
Forward voltage drop	(1mA, 25°C)	0·68V
Forward voltage drop	(10mA, 25°C)	0·87V
Forward voltage drop	(30mA, 25°C)	1·16V
Capacitance	(−0·75V)	0·75pF
Total recovered charge	(+10mA to −5V)	3pC

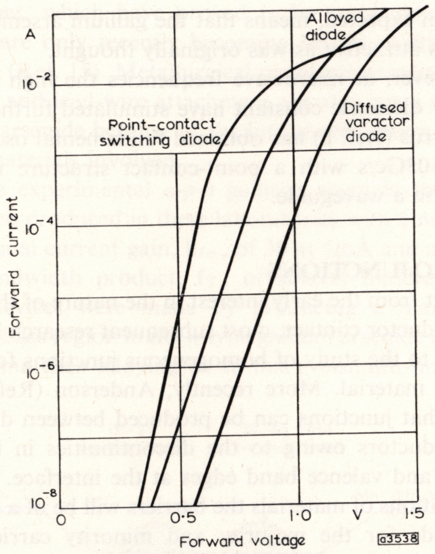


Fig. 3—Forward characteristics of gallium arsenide diodes

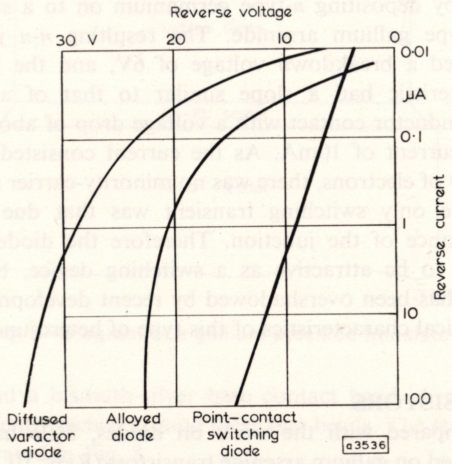


Fig. 4—Reverse characteristics of gallium arsenide diodes

An experimental gallium arsenide varactor diode for operation at low temperatures has been developed at the Wembley Laboratories of Associated Semiconductor Manufacturers Ltd. and is shown in Fig. 5. This diode utilises a diffused mesa junction which is designed to provide optimum values of spreading resistance and capacitance. The characteristics are given in Figs. 3 and 4, and measurements at 10Gc/s indicate that cut-off frequencies of the order of 250Gc/s have been achieved.

By reducing the junction area still further it is possible to improve the frequency performance, and a point-contact device becomes attractive. Although this reduces the permissible dissipation, cut-off frequencies of 800Gc/s have been reported (Ref. 5) with a point-contact diode on heavily doped material and mounted directly in a waveguide.

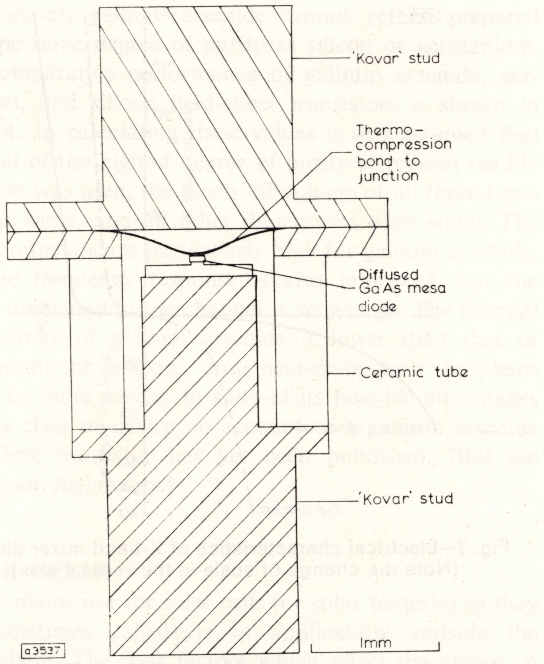


Fig. 5—Diagram of a diffused varactor diode

MICROWAVE MIXER DIODES

As with the higher-frequency varactor diodes, a point-contact junction is employed to obtain the low capacitance required for microwave mixer diodes. For this form of junction the performance depends on similar parameters to the parametric amplifier varactor, and therefore gallium arsenide should offer the best combination of mobility and dielectric constant. Fig. 6 shows the construction of X- and Q-band devices assembled to assess the performance of gallium arsenide in this application. These diodes were produced by pulsing a fine phosphor-bronze wire to form junctions on *n*-type material having a mobility of 3000cm^2/Vsec. By comparison, the germanium normally used for mixer diodes has a mobility of about 800cm^2/Vsec. The relative electrical characteristics are shown in Fig. 7. The contribution of noise from the reverse current of a mixer diode is much larger than that due to the forward characteristic, and therefore a small positive bias is beneficial. The optimum bias of 0·4V is larger than the 0·15V employed with germanium mixer diodes because of

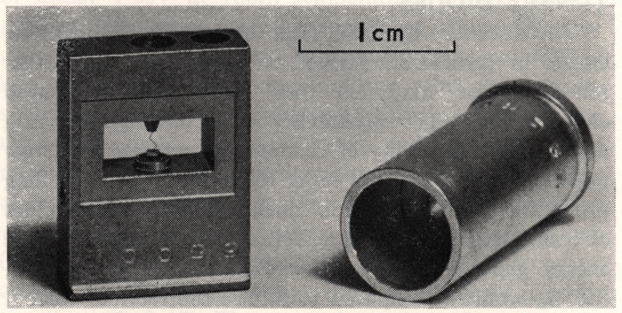


Fig. 6—X- and Q-band mixer diodes

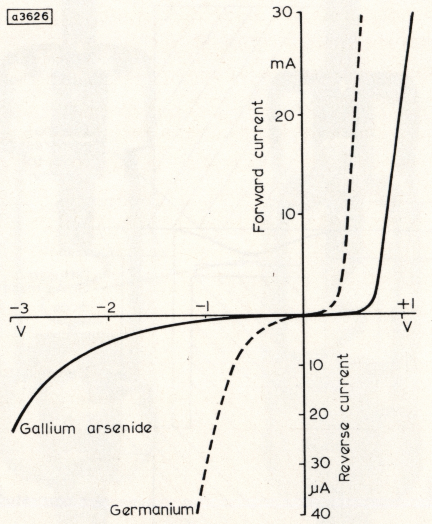


Fig. 7—Electrical characteristics of X-band mixer diodes
(Note the change of scale in the vertical axis)

the larger barrier potential. Under this condition the best noise figure achieved at X-band was 7·3dB in conjunction with a 2dB i.f., and with a smaller-area junction at Q-band a noise figure of 8·1dB was obtained with a 1·5dB i.f. Although these results are comparable with those on commercially available germanium diodes, the extensive development effort on the latter has recently produced superior devices, and further progress with the gallium arsenide diodes awaits improvements in technology.

TUNNEL DIODES

The discovery by Esaki (Ref. 6) in 1958 of tunnelling effects in germanium started a rapid search for other semiconductor materials that would give enhanced characteristics. The probability of tunnelling occurring depends on the ability of carriers to penetrate a narrow potential barrier, and in order to have a high tunnelling probability, a material should have relatively small values of energy gap and carrier effective mass. However, if the energy gap were much lower than that of germanium, cooling would be required to reduce the thermal current level. Of the materials with reasonable energy gaps, gallium arsenide offers a similar tunnelling probability to germanium, and has the advantage of providing an increased power output due to the larger potential barrier.

It was for this reason that a large number of manufacturers developed diodes of gallium arsenide in preference to other materials, and they were therefore dismayed when it was reported that the forward tunnelling characteristic deteriorated irreversibly during the operation of the device.

This degradation has been correlated with a widening of the space-charge region, and is thought to be caused by field-enhanced diffusion of the zinc *p*-type impurity used to form the junction (Ref. 7). This occurs only when the device is operated in the forward injection region. This unfortunate effect, in conjunction with the realisation that the tunnel diode was less useful at low frequencies than had been expected, means that the gallium arsenide diode is not as attractive as was originally thought.

However, at microwave frequencies the high mobility and low dielectric constant have stimulated further work, and Burrus (Ref. 8) has obtained fundamental oscillations up to 103Gc/s with a point-contact structure mounted directly in a waveguide.

HETEROJUNCTIONS

Apart from the early interest in the nature of the metal-semiconductor contact, most subsequent research has been devoted to the study of homogeneous junctions formed in a single material. More recently, Anderson (Ref. 9) has shown that junctions can be produced between dissimilar semiconductors owing to the discontinuities in the conduction and valence band edges at the interface. In most combinations of materials the barriers will be of a different magnitude for the majority and minority carriers, and therefore the current will be mainly conveyed by one type of carrier. In the original devices a heterojunction was made by depositing *n*-type germanium on to a substrate of *n*-type gallium arsenide. The resulting *n-n* junction exhibited a breakdown voltage of 6V, and the forward characteristic had a slope similar to that of a metal-semiconductor contact with a voltage drop of about 0·4V for a current of 10mA. As the current consisted almost entirely of electrons, there was no minority-carrier storage, and the only switching transient was that due to the capacitance of the junction. Therefore the diode would appear to be attractive as a switching device, but that aspect has been overshadowed by recent developments in the optical characteristics of this type of heterojunction.

TRANSISTORS

Compared with the work on diodes, little has been published on gallium arsenide transistors (Refs. 10, 11, 12). In many ways gallium arsenide is an ideal material for transistor use because of its physical properties (Table 1): the wide energy gap should enable devices to be used up to a temperature of approximately 400°C. A material figure of merit for bipolar transistors proportional to $\mu_n\mu_p/\sqrt{\varepsilon}$ was originally proposed by Giacoletto (Ref. 13). On this basis, for reasonable transistor doping levels, gallium arsenide should be almost as good as germanium, and about an order of magnitude better than silicon. Because of the very low minority-carrier lifetime in gallium arsenide the base width of a transistor would have to be very small; although, following Mayburg's analysis (Ref. 14) it should be possible to produce a transistor with a current gain, h_{fe}, of 100 with a base width of 2μm and an average *p*-type impurity density of $5\times10^{17}\text{cm}^{-3}$ in the base. The low lifetime should also make the transistor relatively insensitive to high-energy radiation.

Unfortunately, the high promise offered by the material has not yet been fully realised. Impurity diffusion into gallium arsenide has not proved to be so well-behaved as it is in silicon and germanium, so that the shape of the impurity-concentration profile in the transistor base region is more difficult to control. The oxide-masking planar techniques, which have proved to be so successful with silicon, are only recently becoming feasible with gallium arsenide (Ref. 15). Metallurgical problems associated with alloying and lead-wire attachment are also more severe in gallium arsenide because of the intractable nature of some of the materials involved.

Some experimental *n-p-n* gallium arsenide transistors have been produced in these laboratories with a maximum small-signal current gain, h_{fe}, of 30 at 5mA and a current gain bandwidth product, f_T, of several hundred Mc/s. These devices were made by producing a manganese-diffused base region in the *n*-type gallium arsenide collector region, and then alloying on to this a bismuth-tin emitter bead and a bismuth-silver base contact bead. A mesa is subsequently etched around these two beads. The structure is illustrated in Fig. 8.

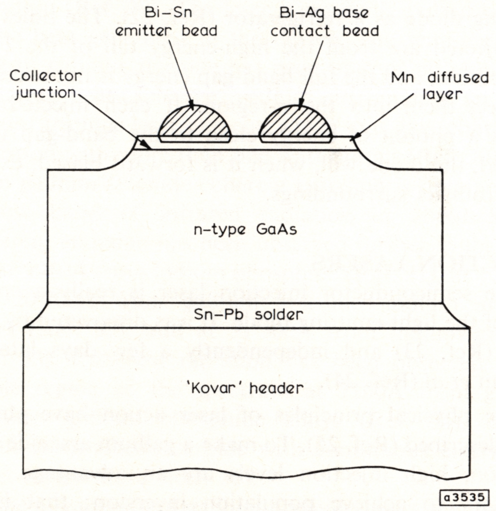


Fig. 8—Diagram of a gallium arsenide transistor

Gallium arsenide also has attractive properties for a field-effect transistor with an *n*-type channel. The high electron mobility is particularly advantageous, but unfortunately gallium arsenide cannot yet be prepared with the same degree of purity as silicon or germanium. The comparative performance of gallium arsenide, germanium, and silicon field-effect transistors is shown in Table 4. In calculating these values it was assumed that material of the highest degree of purity at present readily available was used, the pinch-off voltage of all three types was the same, and all other dimensions were equal. The transconductance is particularly high for gallium arsenide, and the frequency response is also improved, but the power dissipated in the channel is very large. The thermal conductivity of gallium arsenide is lower than that of germanium or silicon, and heat-dissipation problems would be more severe. In spite of its possible advantages for this class of device no account of a gallium arsenide field-effect transistor has yet been published. [But see *Additional References*.]

TABLE 4
Comparison of some Properties of Field-effect Transistors made with Different Semiconductor Materials

Material	V_O	C	f	g_{mO}	I_{DO}	P
Si	1·0	1·0	1·0	1·0	1·0	1·0
Ge	1·0	1·35	2·8	3·8	3·75	3·7
GaAs	1·0	9·4	5·3	50	54	58

V_O = pinch-off voltage
C = gate capacitance
f = limiting frequency
g_{mO} = maximum transconductance
I_{DO} = maximum drain current
P = power dissipation

SOLAR CELLS

The major use for solar cells (or solar batteries as they are sometimes called) is in applications outside the atmosphere. The four factors which affect the choice of material are therefore:

(i) the maximum conversion efficiency of sunlight into electrical energy,
(ii) the temperature performance,
(iii) the resistance to high-energy radiation, and
(iv) the state of technological development which has been reached.

The conversion efficiency depends on the energy gap in the semiconductor material. It has been shown (Ref. 16) that the maximum efficiency is obtained at an energy gap of about 1·5eV at room temperature. In this respect gallium arsenide is more efficient than either silicon or germanium. A choice must be made between materials of low energy gap, where most of the solar photons are sufficiently energetic to produce photoelectrons, so that the short-circuit current is large, and high-energy-gap materials where the open-circuit voltage is high. The absorption coefficient for photons should be high for wavelengths right up to those equivalent to the energy gap (that is, the material should have a steep absorption edge), and in this respect gallium arsenide is markedly superior to silicon (Ref. 17). The steep absorption edge will mean that most of the photons will be absorbed in a region very close to the surface; the *p-n* junction must therefore be very close to the surface. Since the thin layer of material on one side of the *p-n* junction will have a high sheet resistance, it may be better to use an *n*-type layer on a *p*-type substrate rather than the converse, which is more usual. Heavy doping, low-resistance ohmic contacts, and a grid structure to minimise the spreading resistance in the surface layer are also helpful.

The temperature at which a solar cell operates determines its power output and efficiency: as the temperature increases the efficiency decreases. The theoretical efficiency of gallium arsenide decreases from 29 per cent at 0°C to 3 per cent at 300°C, which is an improvement on silicon which decreases from 24 per cent at 0°C to zero at 250°C (Ref. 16). This effect will be particularly important where it is necessary to concentrate the solar energy on to the cell, to increase the power output. If the temperature can be kept low and the impurity concentration in the material is high, the potential barrier height of the junction is nearly equal to the energy gap in the material, and the cell voltage will be a maximum.

Exposure to high-energy radiation will reduce the efficiency of a solar cell; the ultimate criterion for performance is the power the cell is able to deliver after a substantial exposure to this radiation. The degradation is mainly due to a reduction in minority-carrier lifetime. In the case of gallium arsenide the minority-carrier lifetime is already very low, of the order of 10^{-10}s (Ref. 18); consequently a large number of defects may be introduced before the lifetime is affected, and the resistance to high-energy radiation is correspondingly high. The higher resistance of gallium arsenide solar cells to high-energy radiation damage is also due to the fact that the active region of the cell is close to the surface, where only a small amount of this radiation is absorbed. These cells are, however, more susceptible to damage by low-energy protons than the silicon cells (Ref. 19).

The solar cell should be thin with respect to the diffusion lengths of the minority carriers (that is, the average distance through which they diffuse before recombining) but long enough so that all of the incident radiation will be absorbed. The steep absorption edge and low minority-carrier lifetime of gallium arsenide may enable thin-film devices to be made, with a consequent reduction in weight and, possibly, material cost. The surface reflectivity should be minimised by the use of transparent coatings with appropriate thickness and refractive index.

In general terms, a good gallium arsenide solar cell should give a slightly higher short-circuit current than a silicon unit, and about 50 per cent greater voltage. The efficiency of practical silicon cells is unlikely to exceed 15 per cent even with intensive development, whereas gallium arsenide cells may be possible with efficiencies ultimately approaching 25 per cent (Ref. 17).

LIGHT-EMITTING DIODES

Optical effects in gallium arsenide have recently become of great interest. The basic mechanism is the radiation emitted by the recombination of electrons or holes injected over a forward-biased *p-n* junction. This process is called injection electroluminescence. The wavelength of the radiation emitted depends on the energy gap of the semiconductor if recombination occurs via a direct band-gap transition.

In the case of gallium arsenide it has been reported that most of the radiation comes from the *p*-region, indicating that an acceptor centre is involved in the radiative recombination process (Ref. 20) although the exact radiative transition involved is still uncertain. The wavelength is about 8450Å and is therefore in the infrared region of the spectrum. Some semiconductors which emit visible recombination radiation are silicon carbide, gallium phosphide, and some gallium phosphide-gallium arsenide alloys.

Gallium arsenide light-emitting diodes can be used as optical links for switching between circuits in a computer. A suitable detector would be a silicon diode or a gallium arsenide-germanium heterojunction (Ref. 21). No electrical connection is needed between the circuits, and this can be an advantage if the circuits are at different potentials, or if the length of wire required would reduce the speed of response. Other information, such as radio or television signals, can also be transmitted in this way.

The advantages of these light sources lie in their speed of response and ease of modulation. A disadvantage is that the beam can be interrupted or scattered by material in the line of sight.

One unusual application may be the use of a light-emitting diode as a refrigerator (Ref. 22). The holes that are injected are from the high-energy tail of the Fermi distribution, and the full band-gap energy is not necessary to inject them into the *n*-region. If each injected hole causes a photon of energy close to the band-gap to be emitted, the diode will, when it is forward biased, extract heat from its surroundings.

INJECTION LASERS

The semiconductor injection laser is really a special case of the light-emitting diode. It was discovered by Hall *et al* (Ref. 23) and independently a few days later by Nathan *et al* (Ref. 24).

The physical principles of laser action have already been described (Ref. 25). To make a gallium arsenide laser function, high injection levels are an advantage. It is necessary to achieve population inversion; that is, to have a large number of carriers at a high energy level so that, when recombination takes place, they will emit light. Other advantageous conditions are that the temperature should be low, the material should be heavily doped, and it should possess a direct band gap so that phonon co-operation is not necessary in the recombination process. These factors are interdependent; for instance at a high enough current density, with suitable energy levels, it is possible to make a laser that operates at room temperature. If the junction is operated at temperatures below 100°K, however, the conversion efficiency for changing electricity into light can approach unity.

A necessary condition is that the gallium arsenide crystal be shaped into the form of an optical resonator. If the photons move only across the junction its width is too small to act as a resonant cavity. If the photons can be permitted to move along the junction, using the junction as a light guide (Ref. 26), oscillation becomes a possibility and the device can act as a laser. To achieve this the junction itself must be very flat; hence the roughness of the original surface must not be greater than the width of the light guide. Two of the end surfaces of the gallium arsenide are polished orthogonal to the plane of the junction, and these play the part of the mirrors in the gas laser. They need not be silvered, because the reflectivity of gallium arsenide is nearly 30 per cent. The wavelength of the radiation emitted corresponds to one of the cavity modes of the resonator. The beam spread will depend on the parallelism of the reflecting ends (Ref. 27); also, because the emission is from a thin active region, it is diffracted as though it were passing through a narrow slit.

When stimulated emission of the radiation occurs, the spectral output narrows from a band about 100Å wide to one only about 0·5Å wide initially; at higher current levels satellite lines develop with a total band spread of 20Å. The radiation output increases several-fold, and a high percentage of the beam becomes polarised and in phase; that is, coherent. This is because the stimulated photons are exactly similar to the stimulating ones. Interference patterns can be detected.

Room-temperature operation of a gallium arsenide laser has been obtained (Ref. 28), and also continuous operation (Ref. 29). The laser is normally pulsed because of the heat generated in the device. The semi-insulating form of gallium arsenide has been used as an electrically insulating heat-exchange medium surrounding an active gallium arsenide laser element, which thus gave continuous operation with more than one watt output (Ref. 30).

In gallium arsenide coherent emission is obtained as a narrow band of infrared radiation at about 8400Å. Coherent emission has been observed further in the infrared at 9060Å from indium phosphide (Ref. 31), at 17 700Å and 20 700Å from gallium arsenide—indium arsenide alloys (Ref. 32), and at 31 500Å from indium arsenide (Ref. 33). Coherent radiation has been observed in the visible region of the spectrum at 7100Å from gallium arsenide—gallium phosphide alloys (Ref. 34), and at 4560Å from silicon carbide with continuous emission at room temperature (Ref. 35).

The advantages of gallium arsenide lasers are that they can be driven directly. They can be modulated at frequencies greater than 1Gc/s, and the radiation is produced at a high efficiency.

OPTOTRANSISTORS

The gallium arsenide optotransistor, or beam-of-light transistor, was first published by Rediker *et al* (Ref. 21). The original proposal was to use a light-emitting gallium arsenide *p-n* junction for the emitter and an *n*-type gallium arsenide—*p*-type germanium heterojunction for the collector.

The main advantage of the optotransistor is that it enables a semiconductor with a very low minority-carrier lifetime (in fact it is important that the lifetime should be low) to be made into a transistor with a relatively large base width. Other advantages are that the base resistance can be made very small (of the order of 1Ω is considered possible) so that the maximum stable power gain in the common-base connection should be much higher for the optotransistor than the bipolar transistor. The transport from emitter to collector is at the speed of light in the material. The collector efficiency should be high, because essentially all of the infrared radiation incident on the collector heterojunction produces carriers within one micrometre of this junction.

The low carrier lifetime is necessary because the time taken for the carriers to recombine (hence the delay before they emit light) is the most important limiting time-constant of the device. A low lifetime will also give a low emitter diffusion capacitance. The small base resistance arises from the fact that carrier injection can be from the base into the emitter (since it does not matter on which side of the *p-n* junction the light is generated); hence the base resistivity may be decreased and the emitter resistivity increased. In fact it may be more efficient to inject into the *p*-side of the junction since most of the radiation may be generated there (Ref. 20). The higher emitter resistivity can be used to reduce the emitter transition capacitance, and the geometry of the structure will prevent the emitter series resistance from becoming troublesome. The resistivities of the germanium and the gallium arsenide at the collector junction must be adjusted so that the photo-produced carriers can overcome the barrier at the heterojunction, without band-to-band tunnelling or avalanche multiplication taking place.

In order to collect a large percentage of the emitted radiation at the collector, the transistor surfaces may be covered with reflecting coatings to reflect the light towards the collector. It may be possible to produce a coherent mode at the emitter for which all the radiation is aimed at an appropriately placed collector (Ref. 21).

A gallium arsenide *p-n-p* optotransistor with a heterojunction collector has since been realised by Fang *et al* (Ref. 36). This early device had a *p*-type epitaxially-grown germanium collector doped to a level of about 10^{17}cm^{-3}, and hence a space-charge layer width of only about one tenth of a micrometre. The collector efficiency was not therefore very high. This device had a base width of 175μm and a small-signal common-base current gain, h_{fb}, of less than 0·01 at 77°K. A gallium arsenide *p-n-p* optotransistor with a homojunction collector was made by Rutz (Ref. 37) using zinc diffusion into tellurium-doped gallium arsenide. This device had a base width of 50μm and a rather better h_{fb} of 0·24 at 77°K. It was thought that the current gain could be improved by adding reflecting coatings to the device.

The gallium arsenide heterojunction-collector optotransistor has been analysed by Foyt (Ref. 38), who assumed the recombination time to be negligible. He considered coherent and incoherent light emission from the emitter, and compared the results with those of a conventional u.h.f. mesa transistor with equivalent geometry, except for a 0·25μm base width. The advantages of the optotransistor were a lower noise figure and a slightly higher current-gain bandwidth product, f_T, for the coherent mode, also a much greater maximum stable power gain for both modes. Gallium arsenide hetero- and homojunction optotransistors were analysed by Yu (Ref. 39), who took account of the finite lifetime. For gallium arsenide a typical minority-carrier lifetime could be of the order 10^{-9} to 10^{-10}s. If (as it is reasonable to assume) all the other time-constants in the transit-time expression can be made small compared with this, then the f_T of the heterojunction-collector optotransistor would be double that of the homojunction type. In order to have the frequency response of an optotransistor comparable with, or better than, a bipolar transistor, with base width of 0·25μm, the carrier lifetime would have to be less than 3×10^{-12}s. The carrier lifetime is, therefore, a major limiting factor in the frequency response of optotransistors.

CONCLUSION

Some solid-state devices have been considered, and it is clear that gallium arsenide offers very attractive possibilities for most of them. In particular, the new devices using injection electroluminescence have recently aroused great interest. The uses of the semi-insulating form of gallium arsenide are as yet largely unexplored; in particular, its use as a substrate on which to build solid circuit elements is an attractive proposition.

The main obstacles to progress at the present time are the difficulty of preparing high-quality material and the technological difficulties involved in using this material to produce solid-state devices. It is expected that, as further progress is made in this field, very-high-quality gallium arsenide devices will become available.

REFERENCES

1. Welker, H. **Z. Natwf.,** 7, 1952, 744.
2. Dale, J. R., and Josh, M. J. **Solid State Electronics,** 7, 1964, 177.
3. Krakauer, S. M. **Proc. I.R.E.,** 50, 1962, 1665.
4. Hanson, D. C. *et al.* International Solid State Circuits Conf., University of Pennsylvania, 1963. Digest of Papers, p. 54.
5. **Electronics,** 36, 1963, 31.
6. Esaki, L. **Phys. Rev.,** 109, 1958, 603.
7. Longini, R. L. **Solid State Electronics,** 5, 1962, 127.
8. Burrus, C. A. **J. Appl. Phys.,** 32, 1961, 1031.
9. Anderson, R. L. **Solid State Electronics,** 5, 1962, 341.
10. Jenny, D. A. **Proc. I.R.E.,** 46, 1958, 959.
11. Jones, M. E. and Wurst, Jnr., E. C. **I.R.E. Int. Conv. Record ED,** 9, 1961, 26.
12. Haidemenakis, E. D. *et al.* **Proc. I.R.E.,** 49, 1961, 1448.
13. Giacoletto, L. J. **R.C.A. Rev.,** 16, 1955, 34.
14. Mayburg, S. **Solid State Electronics,** 2, 1961, 195.
15. Shortes, S. R. *et al*, Ing, S. W. *et al* and Hall, M. A. *et al.* Electrochem. Soc. Meeting, Pittsburgh, April 15-18, 1963.
16. Wysocki, J. J. and Rappaport, P. **J. Appl. Phys.,** 31, 1960, 571.
17. Moss, T. S. **Solid State Electronics,** 2, 1961, 222.
18. Loferski, J. J. and Wysocki, J. J. **R.C.A. Rev.,** 22, 1961, 38.
19. Wysocki, J. J. **J. Appl. Phys.,** 34, 1963, 2915.
20. Nathan, M. I. and Burns, G. **Appl. Phys. Letters,** 1, 1962, 89.
21. Rediker, R. H. *et al.* **Proc. I.E.E.E.,** 51, 1963, 218.
22. Keyes, R. J. and Quist, T. M. **Proc. I.R.E.,** 50, 1962, 1822.
23. Hall, R. N. *et al.* **Phys. Rev. Letters,** 9, 1962, 366.
24. Nathan, M. I. *et al.* **Appl. Phys. Letters,** 1, 1962, 62.
25. Hayward, R. J. R. **G.E.C. Journal,** 29, (2) 1962, 92.
26. Ashkin, A. and Gershenzon, M. **J. Appl. Phys.,** 34, 1963, 2116.
27. Burns, G. *et al.* **I.B.M. Jnl. Res. and Dev.,** 7, 1963, 62.
28. Burns, G. and Nathan, M. I. Ibid p. 72.
29. Howard, W. E. *et al.* Ibid p. 74.
30. Garfinkel, M. and Engeler, W. E. I.E.E.E. Solid State Device Research Conf., East Lansing, Michigan, June, 1963.
31. Weiser, K. and Levitt, R. S. **Appl. Phys. Letters,** 2, 1963, 178.
32. Melngailis, I. *et al.* **Proc. I.E.E.E.,** 51, 1963, 1154.
33. Melngailis, I. **Appl. Phys. Letters,** 2, 1963, 176.
34. Holonyak, N. and Bevacqua, S. F. **Appl. Phys. Letters,** 1, 1962, 82.
35. Griffiths, L. B. *et al.* **Proc. I.E.E.E.,** 51, 1963, 1374.
36. Fang, F. F. *et al.* Ibid p. 860.
37. Rutz, R. F. Ibid p. 470.
38. Foyt, A. G. Ibid p. 853.
39. Yu, H. N. Ibid p. 945.

ADDITIONAL REFERENCES

GaAs insulated-gate field-effect transistors have now been described in the two following references.

Becke, H., *et al.* U.S. Govt. Rep. AD602440.
Charlson, E. J. U.S. Govt. Rep. AD603752.

Light-operated solid-state devices ('transluxors') incorporating heterojunctions of GaAs and another semiconductor material were exhibited by Mullard Research Laboratories at the Institute of Physics and Physical Society Exhibition, January 1964.

Transient Performance of Heatsinks for Power Devices

J. H. Tuley
Mullard Semiconductor Electrical Development Laboratory

The thermal performance of heatsinks during temporary current overloads can be calculated simply from the thermal capacity of the heatsink, or more accurately from the thermal capacity and the thermal time-constant. The simple calculation gives pessimistic results, and heatsinks designed to allow for the thermal effects calculated in this way would have a substantial margin of safety.

The expressions given can be used to calculate the maximum tolerable duration of an overload, or the maximum allowable initial current where an overload of known amplitude and duration will be superimposed, or the maximum tolerable overload that may be superimposed on a given initial current for a given time.

One or other of these conditions frequently has to be solved when thyristors or other power devices are used with loads, such as motors, which draw variable currents.

INTRODUCTION

The purpose of a heatsink is normally regarded as the provision of a suitable 'thermal resistance', which allows a power device to be run continuously at a given dissipation without a specified value of mounting base temperature being exceeded. A heatsink also provides a useful 'thermal capacity' which allows appreciable overload currents to be carried for periods of up to a few minutes.

To obtain a given thermal resistance with the minimum quantity of heatsink material requires a design with relatively thin fins. On the other hand, to obtain a large thermal capacity for a given thermal resistance requires the part of the heatsink on which the device is mounted to be relatively thick.

A compromise has to be achieved between these two requirements, and this compromise is often linked with mechanical considerations. In the case of the 30D, 40D, and 50D extrusions (Fig. 1) the centre part of the extrusion was limited in thickness by the requirement to fix studded devices (for example, the BYZ14) through the centre part with a nut on the other side. This type of design resulted in thermal time-constants, for natural convection, of about 5 minutes for the rise in heatsink temperature.

For the 60D extrusion (Fig. 2), designed primarily for the flat-based BYX14 device, the centre part is considerably thicker. This means that when a studded device is used, the hole for mounting has to be specially drilled and tapped. However, the thermal time-constant for the 60D extrusion is about twice that of the other extrusions, resulting in an improved transient performance.

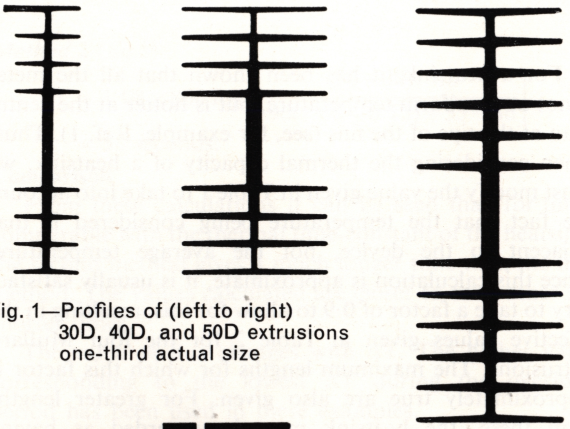


Fig. 1—Profiles of (left to right) 30D, 40D, and 50D extrusions one-third actual size

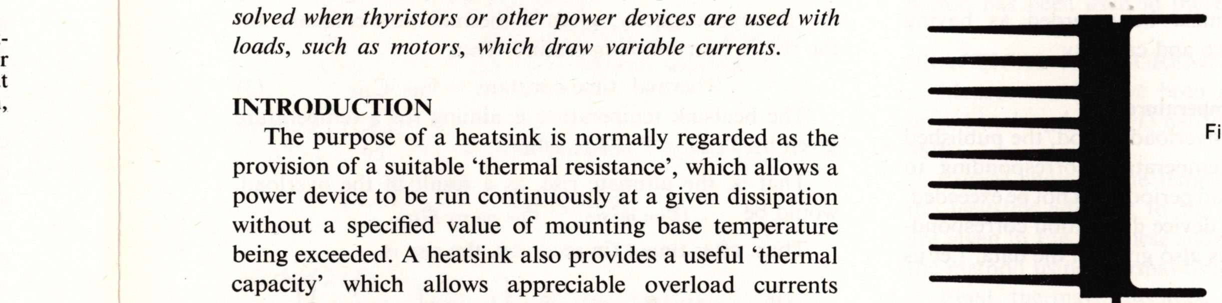


Fig. 2—Profile of 60D extrusion one-third actual size

In some installations forced-air cooling can be used with advantage to reduce the thermal resistance of a given heatsink. An unfortunate side-effect is that the thermal time-constant is also reduced. In consequence, special care must be taken when considering overload currents with forced-air cooling.

THEORETICAL CONSIDERATIONS

Two approaches can be used when considering the transient performance of 30D, 40D, 50D, and 60D heatsinks. These are a simple method, which uses only the thermal capacity of these extrusions, and another which uses the thermal time-constant.

Although either of these methods can be used for these types of extrusion, based on the same properties of the material, they cannot always be applied to other types of heatsink, for example thin plates.

Properties of Heatsink Materials

The thermal capacities of the various metals normally used for heatsinks are given in Table 1.

TABLE 1

Thermal Capacities of Heatsink Materials

Metal	**Thermal Capacity** (watt.sec/lb.degC)
Aluminium	407
Copper	175
Steel	204

TABLE 2

Effective Thermal Capacities of Aluminium Extrusions

Extrusion	**Weight** (lb/ft)	**Effective Thermal Capacity/inch** (watt.sec/degC)	**Max length** (in.)
30D	1·1	34	5
40D	1·8	55	6
50D	2·7	83	6
60D	7·1*	215	8

**for calculation of thermal capacity; the actual weight may be higher because of a recent change of profile.*

For a heatsink, it has been shown that all the metal is not at a uniform temperature, but is hotter at the centre than at the tips of the fins (see, for example, Ref. 1). Thus, when considering the thermal capacity of a heatsink, we must modify the value given in Table 1 to take into account the fact that the temperature being considered is that adjacent to the device, not the average temperature. Since this calculation is approximate, it is usually satisfactory to take a factor of 0·9 to allow for this, resulting in the effective values given in Table 2 for the four Mullard extrusions. The maximum lengths for which this factor is approximately true are also given. For greater lengths than these, the heatsink must be regarded as having distributed thermal resistance and capacity.

Equations for Operating Temperatures

At the conclusion of the overload period, the published limit for mounting base temperature corresponding to the current during the overload period must not be exceeded. Suppose this is $T_{mb(o)}$. The device dissipation corresponding to the overload current is also given in the data. Let us call this $P_{tot.max(o)}$.

If we take into account the contact thermal resistance for the device, θ_i, the maximum allowable heatsink temperature is then

$$T_{mb(o)} - P_{tot.max(o)}\theta_i.$$

If the device has initially been carrying a current continuously before the overload, then, from the data, this gives a device dissipation of $P_{tot.max(i)}$.

The initial heatsink temperature is given by

$$T_{amb} + P_{tot.max(i)}\theta_{HS}$$

where θ_{HS} is the thermal resistance of the heatsink.

The allowable rise in heatsink temperature is then

$$\Delta T_{HS} = T_{mb(o)} - P_{tot.max(o)}\theta_i - T_{amb} - P_{tot.max(i)}\theta_{HS}. \qquad \ldots(1)$$

If there is no initial current, this equation can be simplified by deleting the last term.

If this expression gives a negative value of ΔT_{HS}, then the overload with which it is proposed to load the device is too great. Either the overload requirement must be reduced, or a larger heatsink used, or a higher current rated device used.

Simple Calculation of Transient Performance

The simple method used in determining the transient performance is to consider only the thermal capacity of the heatsink. The time for which the overload current can be carried is then given by

$$t \simeq \frac{C_{HS} \times \Delta T_{HS}}{P_{tot.max(o)}} \text{seconds} \qquad \ldots(2)$$

where C_{HS} = thermal capacity of heatsink
= (effective thermal capacity/inch) ×
(length in inches)

Transient Performance using Thermal Time-constant

The thermal resistance of the heatsink, together with the thermal capacity, gives a thermal time-constant for the rise in heatsink temperature. Thus

$$\text{Thermal time-constant} = \theta_{HS}.C_{HS} \qquad \ldots(3)$$

The heatsink temperature is aiming for a temperature given by $T_{amb} + P_{tot.max(o)}\theta_{HS}$.

That is, the ultimate rise, as a result of the overload, would be $(P_{tot.max(o)} - P_{tot.max(i)})\theta_{HS}$.

Thus, after time t (in seconds), the rise is

$$(P_{tot.max(o)} - P_{tot.max(i)})\theta_{HS}\left[1 - \exp\left(\frac{-t}{\theta_{HS}.C_{HS}}\right)\right].$$

Equating this to the allowable rise ΔT_{HS} gives

$$\Delta T_{HS} = (P_{tot.max(o)} - P_{tot.max(i)})\theta_{HS}\left[1 - \exp\left(\frac{-t}{\theta_{HS}.C_{HS}}\right)\right]. \qquad \ldots(4)$$

This can be manipulated to give

$$t = \theta_{HS}C_{HS}.\log_e\left[\frac{(P_{tot.max(o)} - P_{tot.max(i)})\theta_{HS}}{P_{tot.max(o)}(\theta_{HS}+\theta_i) + T_{amb} - T_{mb(o)}}\right]. \qquad \ldots(5)$$

This is the maximum allowable time for which the overload can be carried.

By suitable rearrangement, the solutions to the following problems can also be obtained.

(a) If a power device and heatsink are required to carry a specified overload current for a specified time, what is the maximum allowable initial current before the overload?

(b) What overload can be carried for a specified time following a specified initial current? In this case, both $P_{tot.max(o)}$ and $T_{mb(o)}$ are unknown, therefore successive approximations will be required.

Repeated and Two-stage Overloads

The time-constant for cooling the heatsink after an overload is the same as that for heating, and it can be used in the calculations for repetitive overloads. If the interval between overloads is at least three times the thermal time-constant, then the heatsink can be assumed to have returned to its starting temperature before the occurrence of the next overload.

With two-stage overloads (that is, where one overload period is immediately followed by a second overload period at a different current level, with no intervening period of cooling), a calculation employing the super-position principle can be used.

EXAMPLES

Two examples are calculated, each by the simple method (Eq.2) and the more accurate method (Eq.5).

Example 1

A BTY99 thyristor is mounted on a 4 in. length of 60D extrusion, carrying a steady initial current of 25A, 180° conduction, with T_{amb} = 35°C and with natural convection cooling. For how long can an overload current of 50A be carried?

From the data

$P_{tot.max(i)} = 30W$
$P_{tot.max(o)} = 74W$
$T_{mb(o)} = 95°C$
$\theta_i = 0{\cdot}15 degC/W$
$\theta_{HS} = 1{\cdot}08 degC/W$ (for ΔT = 30degC)
$C_{HS} = 215 \times 4 = 860 watt.sec/degC.$

Method 1 (*Eq*.2)

$$\Delta T_{HS} = 95-(74\times0{\cdot}15)-35-(30\times1{\cdot}08) = 16\text{degC}$$

$$t = \frac{860\times16}{74} = 186 \text{ seconds.}$$

Method 2 (*Eq*.5)

$$t = 1{\cdot}08\times860\log_e\left[\frac{(74-30)1{\cdot}08}{74(1{\cdot}08+0{\cdot}15)+35-95}\right]$$

$$= 395 \text{ seconds.}$$

Example 2

For the same conditions as in the previous example it is required to find for how long an overload current of 60A can be carried.

From the data

$P_{tot.max(i)} = 30W$
$P_{tot.max(o)} = 96W$
$T_{mb(o)} = 86°C$
$\theta_i = 0{\cdot}15 degC/W$
$\theta_{HS} = 1{\cdot}08 degC/W$ (for ΔT = 30degC)
$C_{HS} = 215 \times 4 = 860 watt.sec/degC.$

Method 1 (*Eq*.2)

$$\Delta T_{HS} = 86-(96\times0{\cdot}15)-35-(30\times1{\cdot}08) = 4{\cdot}2\text{degC}$$

$$t = \frac{860\times4{\cdot}2}{96} = 38 \text{ seconds.}$$

Method 2 (*Eq*.5)

$$t = 1{\cdot}08\times860\log_e\left[\frac{(96-30)1{\cdot}08}{96(1{\cdot}08+0{\cdot}15)+35-86}\right]$$

$$= 58 \text{ seconds.}$$

Comparison

From these calculations it can be seen that, for times comparable with the thermal time-constant of the heatsink, the second method gives a much longer time than the first method. The values calculated are closer to each other for shorter times.

The value of θ_{HS} changes with ΔT. The higher value corresponding to the lowest ΔT during the overload period has been used in these examples.

PRACTICAL MEASUREMENTS

Measurements have been made with a BYZ14 device on a 4 in. length of 60D extrusion, with an input power of 48W.

Curves showing the temperature rise against time for the heatsink adjacent to the device and also the fin tip are given in Fig. 3. These show reasonable agreement with the method using thermal time-constants.

Final thermal resistance = 1·0degC/W, giving final rise of 48degC.

Then thermal time-constant = 860 × 1·0 = 860 seconds.

After this time, the rise should be

0·632 × 48 = 30degC (measured 32degC).

Alternatively, for an average ΔT of 25degC approx.,

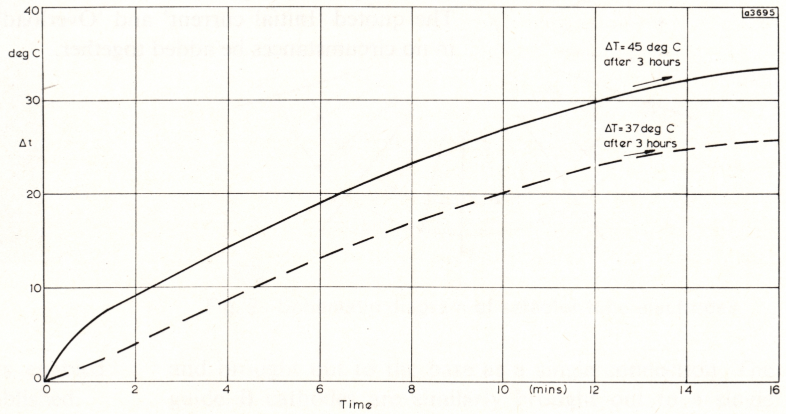


Fig. 3—Temperature gradient of 4in length of 60D unblackened extrusion
Power into centrally-mounted studded device = 48W T_{amb} = 23°C
——————— thermocouple in heatsink at base of device
- - - - - - - thermocouple at tip of outside fin

$\theta_{HS} = 1{\cdot}1$degC/W, which would give a final rise of 53degC. Then, thermal time-constant = 950 seconds.

After this time, the rise should be $0{\cdot}632 \times 53 = 33{\cdot}5$degC, which is closer to the measured value of 34 degC.

It will be noticed that there is a more rapid rise in heatsink temperature adjacent to the device than would be expected from thermal time-constant considerations, during the first minute. This is because the heatsink has distributed thermal resistance and capacity, which have been lumped together to simplify the calculation. It is advisable to allow a fairly generous 'safety factor' when calculating overload conditions for times of one minute or less.

STACK OVERLOADS

There are an infinite number of overload conditions that could be considered. Table 3 gives typical computed maximum overloads for some stacks in the Mullard range of single-phase bridge thyristor stacks.

REFERENCE

1. Tuley, J. H. 'Design of Cooling Fins for Silicon Power Rectifiers'. **Mullard Technical Communications,** Vol. 5, No. 44, June 1960, pp. 118 to 130.

TABLE 3

Computed Transient Performance of Thyristor Stacks

Stack type	**Rated continuous current**	**Overload Conditions**		
		Initial current	**Overload current**	**Overload time (mins)**
OTH17-608	17A	0A 10A	24A 24A	5·0 1·4
OTH23-608	23A	0A 15A 20A	32A 30A 25A	6·5 0·5 3·5
OTH36-608	36A	0A 20A	50A 50A	8·0 1·5
OTH68-608	68A	0A 40A	80A 80A	13·5 4·5
OTH110-608	110A	0A 70A 65A	140A 140A 130A	8·25 0·5 5·5

All currents are for 180° conduction and $T_{amb} = 35$°C

The overload currents are the *total* allowable currents. Thus, in the final example, an initial current of 65A is allowed to have a current of 65A superimposed on it for a maximum time of 5·5 minutes. The maximum allowable current under overload conditions is, in this example, 65+65 = 130A, as given in the 'Overload current' column. The quoted 'Initial current' and 'Overload current' must in no circumstances be added together.

Operating Principles of the Z504S Stepping Tube

C. S. Barker
Mullard Gas Discharge and Photoelectric Devices Division

The characteristics and mode of operation of a conventional ten-position, thirty-cathode stepping tube are described. Three forms of guide pulse drive are given; double rectangular pulse drive; rectangular guide A and integrated guide B pulse drive; differentiated guide A pulse and integrated guide B pulse drive. These waveforms are generally used with stepping tubes and they are discussed in conjunction with the requirements to ensure that the waveforms meet the Z504S characteristic operating areas. Reliability figures obtained from circuits employing practical waveforms are presented. Only one tube failure per one million tube hours has been observed.

INTRODUCTION

The conventional central-anode, thirty-cathode stepping tube has been used successfully for a number of years in circuits requiring visual indication of a counting operation at low cost, combined with low power consumption. An example of this type of tube is the Z504S selector shown in Fig. 1.

This article (which is based in part on material drawn from References 1 and 2) describes the transfer mechanism, the mode of operation and the characteristic working areas for the Z504S selector tube. When considering the working areas the approach has been to study the individual main characteristics and assess the stability of these characteristics, at their natural limits, under a number of life test conditions. These life test experiments enabled reliable working areas for the Z504S to be established.

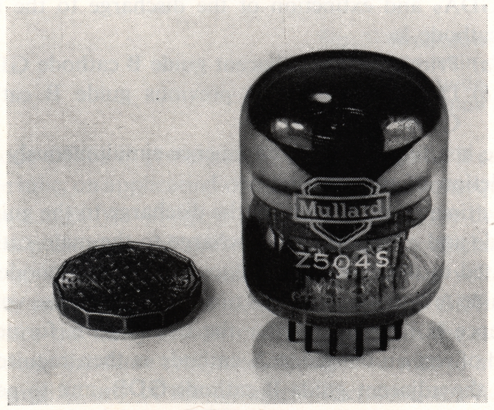


Fig. 1—Z504S selector tube

Stepping tube circuits have the advantage of using only one tube per decade as opposed to a number of active elements as in alternative circuits. Thus, to achieve a similar reliability to that of the Z504S stepping tube, an alternative device would require a considerably lower failure rate for the individual elements in each decade.

The three most common guide waveforms required to transfer the discharge are given. The recommended values are carefully derived from the characteristics discussed later in this article, which ensure that full advantage is taken of the inherent reliability of the Z504S selector tube.

STEPPING TUBE CONSTRUCTION

The Z504S selector tube consists of thirty rod-shaped cathodes which are equally spaced in a circle around a central anode disc. The structure is enclosed in a glass envelope which contains a mixture of inert gases at low pressure.

Fig. 2 shows the connections of the thirty cathodes. They are divided into three interleaved groups of ten, termed main cathode (k_1, k_2, . . .); guide A cathode (GDA_1, GDA_2, . . .); and guide B cathode (GDB_1, GDB_2, . . .). The guide A cathodes are connected together and brought out to the base as a single connection. The guide B cathodes are similarly brought out to a single

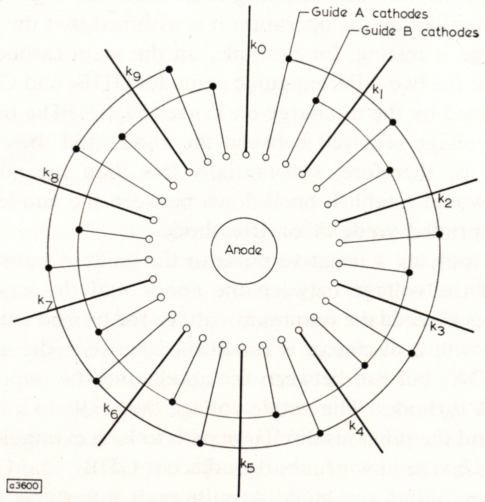


Fig. 2—Schematic diagram of selector tube electrodes

connection, while each main cathode has a separate base connection. Thus an output signal may be taken from any cathode as required. Fig. 3 shows the sequence of cathodes k_7 to k_9 inclusive, which, viewed from left to right, gives the clockwise order of the cathodes around the tube.

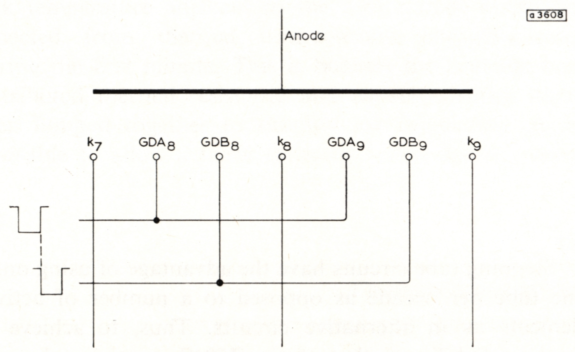


Fig. 3—Sequence of cathodes in the Z504S selector tube

VISUAL INDICATION

In operation, a discharge is set up between the anode and a cathode, and appears as an orange-red glow at the tip of the cathode. With a suitable drive circuit, which is connected to the guide cathodes, the glow is made to travel from one main cathode to the next. The position reached by the glow can be seen through the top of the tube and can be read off from an external escutcheon which bears the numbers 1 to 9 and 0 opposite the appropriate main cathodes. The glow is usually driven clockwise (k_1, k_2, k_3, . . .), but anticlockwise rotation may be used if it is desired. The tube is, in fact, completely bi-directional.

MODE OF OPERATION

The most common method of driving a stepping tube is to apply a pair of negative pulses to the guide cathodes as shown in the basic stepping tube circuit of Fig. 4.

In considering the operation it is assumed that the glow discharge is resting, for example, on the main cathode k_8 and that the two adjacent guide cathodes GDB$_8$ and GDA$_9$ are primed by the discharge on k_8 (see Fig. 3). The breakdown voltage required between the anode and these two guides, is, therefore, substantially less than the voltage which would establish breakdown between the anode and any unprimed guide A or B cathode.

On applying a negative pulse to the guide A cathodes, so that the voltage between the anode and the guide A cathodes exceeds the minimum voltage for primed cathode breakdown, a discharge is established between the anode and GDA$_9$ but not between the anode and the unprimed guide A cathodes. The anode voltage then falls to a lower value and the previous discharge path to k_8 is extinguished. The discharge now primes the adjacent GDB$_9$, and GDB$_8$ deionises. When the guide A pulse ends a negative pulse is applied to the guide B cathodes, a discharge is formed between anode and GDB$_9$, and the GDA$_9$ discharge is extinguished.

At the end of the guide B pulse the guide B cathodes return to the positive guide bias level, and the anode potential rises towards the h.t. supply. When the anode potential reaches a level greater than the anode-to-k_9 breakdown potential (k_9 now being primed by the discharge to GDB$_9$), a discharge is formed to k_9. The anode potential then falls to the maintaining potential of the anode-to-k_9 path, and the discharge to GDB$_9$ is extinguished.

The process of moving the glow discharge from one main cathode to the next has now been completed. It can obviously be repeated for transfer to the next main cathode, and so continue round the tube. For clockwise rotation of the glow, guide A must be pulsed negatively before guide B. Reversal of the order of pulsing will produce anticlockwise rotation.

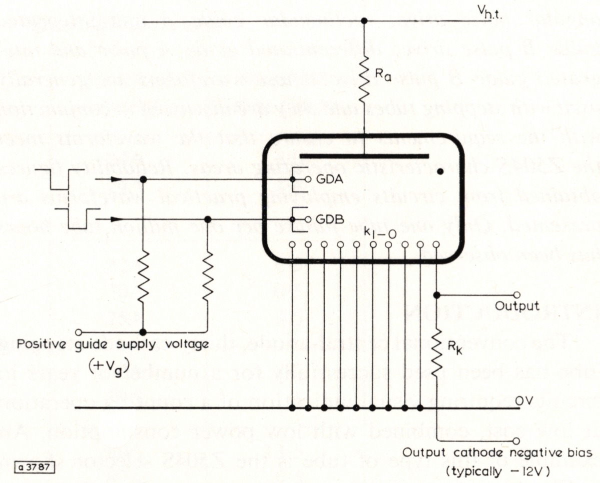


Fig. 4—Basic stepping tube circuit

DETAILS OF TRANSFER MECHANISM

To transfer the discharge from a main cathode, for example, k_8, to the adjacent guide A cathode GDA$_9$, the guide pulse must be applied for a period long enough to allow three processes to take place:

1. Formation of the discharge to the guide A cathode GDA$_9$ and extinction of the discharge to the main cathode k_8.
2. (*a*) Priming of the adjacent guide B cathode GDB$_9$.
   (*b*) Deionisation of the previous guide B cathode GDB$_8$.

The last two processes commence simultaneously after completion of process number 1.

The time required to form the discharge to the guide A cathode and extinguish the discharge to the main cathode is related to the voltage difference between the anode and the guide A cathodes, and this time decreases as the voltage difference is increased. After the discharge is formed, a period of rest on the guide A cathode is then required to prime the adjacent guide B cathode GDB$_9$ and to permit the previous guide B cathode GDB$_8$ to deionise.

Fig. 5(*a*) shows the voltage required for breakdown to each of the guide B cathodes GDB$_8$ and GDB$_9$ as a function of the time for which negative voltage has been applied to the guide A cathodes. Before the crossover point 'X' in the diagram, the breakdown voltage of the unwanted guide GDB$_8$ is less than that of GDB$_9$. After the crossover point, however, the directional priming properties of the tube are set up and breakdown will occur to the guide B cathode GDB$_9$ in preference to any other guide B cathode in the tube, on application of a guide B pulse. Consequently, when a negative voltage is applied to GDA$_9$ the duration of the pulse must be sufficient to allow the discharge to form to GDA$_9$, and to allow the breakdown voltage of GDB$_9$ to fall to a lower value than that of GDB$_8$.

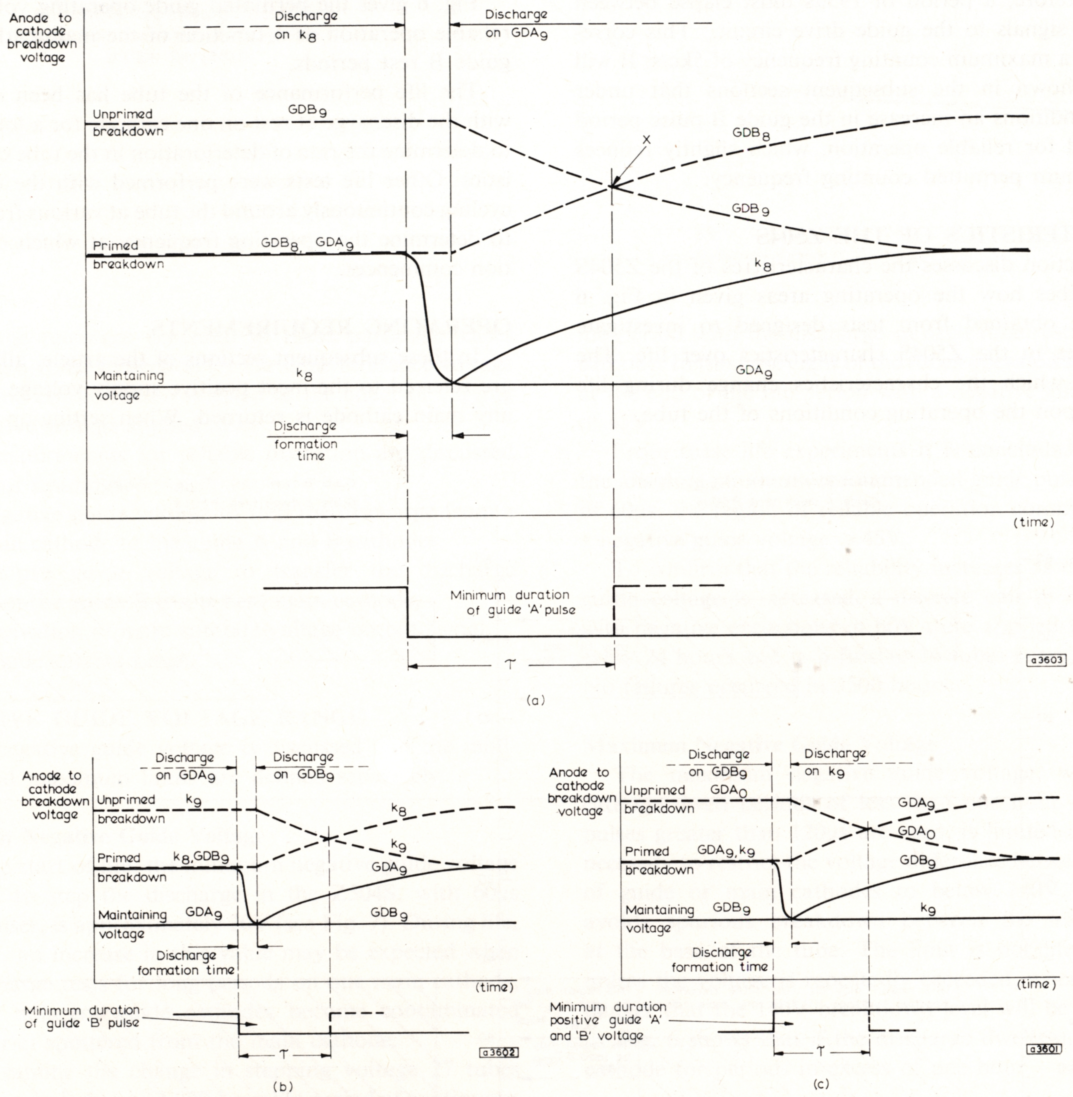


Fig. 5—Breakdown voltage of anode-to-cathode gaps—
(*a*) during main cathode to guide A cathode transfer
(*b*) during guide A to guide B transfer
(*c*) during guide B to main cathode transfer

Since the thirty anode-to-cathode gaps in the tube are identical, the same requirements must be met to step the discharge to any of the cathodes. Therefore, the negative voltage applied to the guide B cathodes to transfer the discharge from GDA$_9$ to GDB$_9$ (Fig. 5*b*) must be applied for a period long enough to ensure:

1. Formation of a discharge to the primed GDB$_9$, and extinction of the discharge on GDA$_9$.
2. (*a*) Priming of the adjacent main cathode k_9.
   (*b*) Deionisation of the preceding main cathode k_8.

As shown earlier in this article (page 132), the discharge is transferred from GDB$_9$ to k_9 when the GDB$_9$ potential approaches the positive supply voltage such that the sum of the positive guide supply voltage and the maintaining voltage of the anode-to-GDB$_9$ gap is greater than the primed breakdown of k_9. The main cathode, k_9, should be at a negative potential with respect to the guide cathodes for a period long enough to ensure that the discharge is formed to k_9, and that the breakdown potential of the succeeding guide A cathode GDA$_0$ can fall to a lower value than that of the preceding guide A cathode GDA$_9$, by the instant when a negative potential is to be applied to the guide A cathodes (Fig. 5*c*).

In the following sections it has been assumed that the minimum period for which a pulse may be applied to the guide electrodes to transfer the discharge from k_8 to GDA$_9$

or from GDA_9 to GDB_9 and to ensure that the correct priming levels are obtained, is 60µs. The minimum time to transfer the discharge from GDB_9 to k_9 and to ensure that the discharge will step to GDA_0 on receipt of the next negative pulse being applied to the guide electrodes, is 75µs. Therefore, a period of 195µs must elapse between successive signals to the guide drive circuit. This corresponds to a maximum counting frequency of 5kc/s. It will also be shown in the subsequent sections that under certain conditions an increase in the guide B pulse period is required for reliable operation, which slightly reduces the maximum permitted counting frequency.

CHARACTERISTICS OF THE Z504S

This section discusses the characteristics of the Z504S and describes how the operating areas given in Fig. 6 have been obtained from tests designed to investigate the changes in the Z504S characteristics over life. The degree to which the characteristics change during life depends upon the operating conditions of the tube.

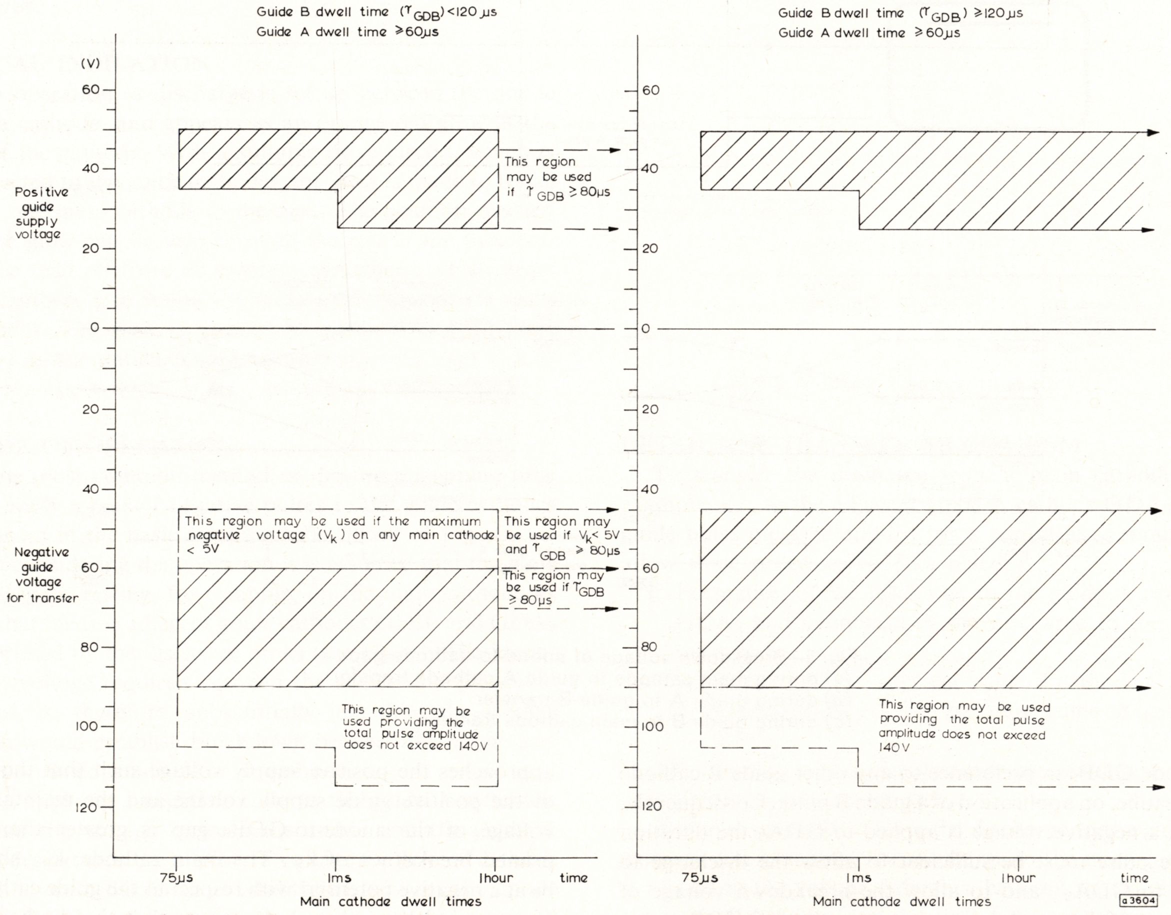


Fig. 6—Guide operating voltages. The shaded areas represent regions where the tube may be used without restriction initially and during life

Variations in maintaining and breakdown voltages which may occur during life tend to reduce the characteristic operating ranges. These changes are generally due to cathode material being sputtered from the cathode on which the discharge rests on to the other cathodes in the tube. This sputtered material has the effect of contaminating these cathodes. The changes are, in general, reversible and when the discharge is moved to a contaminated cathode the sputtering process cleans the cathode surface again.

Fig. 6 gives the permitted guide operating voltages for reliable operation as a function of the main cathode and guide B rest periods.

The life performance of the tube has been examined with the discharge at rest on one cathode for a long period to determine the rate of deterioration in the tube characteristics. Other life tests were performed with the discharge cycling continuously around the tube at various frequencies to determine the operating frequency at which deterioration commences.

OPERATING REQUIREMENTS

In these subsequent sections of the article, all voltages are referred to the most positive supply voltage to which any main cathode is returned. When setting-up life tests and obtaining measurements, for Sections (*a*) and (*b*), the main cathode circuit employed has cathodes 1 to 9 inclusive returned to zero potential and k_0 has a negative voltage of 12V applied to it as shown in Fig. 4. Consequently the reference potential in these particular cases will be zero volts. In Section (*c*) the same reference voltage is used and consideration is given to the effect of varying the negative voltage applied to k_0.

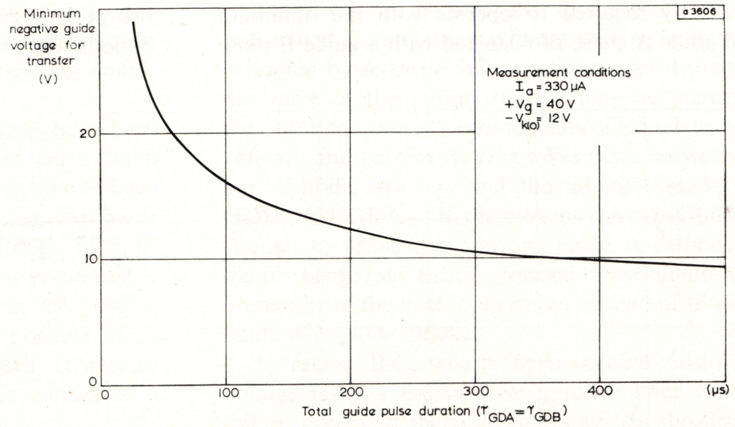


Fig. 7—Typical variation of minimum negative guide voltage for transfer with guide pulse duration

The requirements for reliable operation are discussed under four headings:

(*a*) Negative guide voltage to step the discharge from a main cathode to the guide A and B cathodes.
(*b*) Positive guide voltage to transfer the discharge from the guide B to the next main cathode.
(*c*) Derivation of main cathode voltage output signal.
(*d*) Anode current range.

NEGATIVE GUIDE VOLTAGE RANGE

The negative guide voltage is discussed and the minimum and maximum limits are defined separately.

Minimum Negative Guide Voltage

At the start of life, the minimum negative guide voltage required to step the discharge in the Z504S, with 60µs guide pulses, is approximately 20V (see Fig. 7). During life, however, an increase in this value may be expected when the discharge rests for long periods on one main cathode, and the adjacent guide cathodes become contaminated by material sputtered from the main cathode.

To examine the change in stepping voltage 27 tubes were measured after a 14 000-hour life period of continuous cycling. During this life period, input signals ranging from one every 10ms to one every 276 hours were fed to the tubes. Each tube was operated at a specific frequency within the given range. The results show a gradual increase in minimum negative guide voltage as the operating speed is reduced. In all except one tube, however, a negative guide voltage of 45V was sufficient to step the discharge.

The results and the end of life measurement conditions are given in Fig. 8.

The requirements for stepping the discharge after long periods of rest on one cathode with guide pulses of 60µs duration have been examined in more detail under standby life conditions. In one experiment 20 tubes were operated for 1700 hours during which time the discharge was stepped from one cathode to the next at 85-hour intervals. Each step was made with a negative guide voltage of 40V, and only two tubes failed to step correctly at all times. A group of ten tubes was also measured after a 9000-hour life period with the discharge in each tube resting on one cathode. In this case eight of the tubes operated successfully at the end of the life period with a negative guide voltage of 45V.

From these life experiments it is concluded that with the absolute minimum recommended guide pulse duration of 60µs, the region of reliable operation commences with a negative guide voltage ≥45V.

To confirm that the reliability increases as the negative guide voltage is increased, a discrete pair of 60µs pulses with negative excursions of 60V were applied to 16 tubes every 24 hours and to a further 16 tubes every 168 hours. No failures occurred in 3500 hours.

Maximum Negative Guide Voltage

The maximum negative guide voltage, where main cathode dwell periods of less than one hour or guide B pulses greater than 120µs are used, is limited only by the necessity to restrict the voltage difference between any pair of guide or main cathodes to below 140V. This limit avoids spurious breakdown between the lead-in wires at the base of the tube. The limit is not life-dependent unless the voltage is repeatedly exceeded, when there is a danger that the 140V breakdown level will be reduced.

Fig. 6 shows that if the discharge dwells upon a main cathode for periods in excess of one hour – and the tube

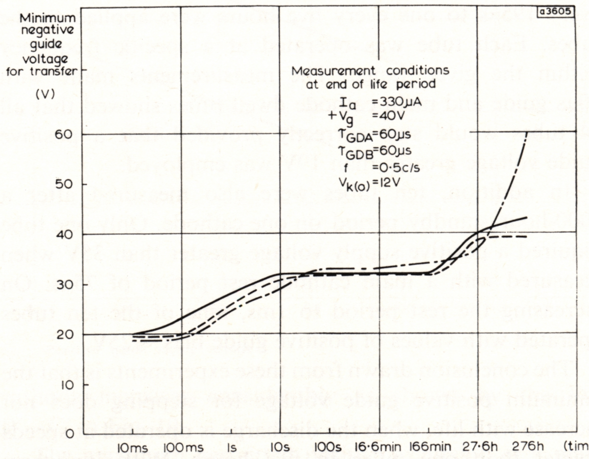


Fig. 8—Minimum negative guide voltage for transfer after a 14 000-hour life period. (The horizontal axis represents the input signal frequency during the life period.)

is subsequently required to operate with the minimum permitted guide A pulse of 60μs and with a guide B pulse between 80μs and 120μs – the maximum negative guide voltage for stepping is reduced to 70V. As the mechanism of failure in this instance for values >70V has to be considered in conjunction with the positive guide supply voltage, it will be described in a separate section.

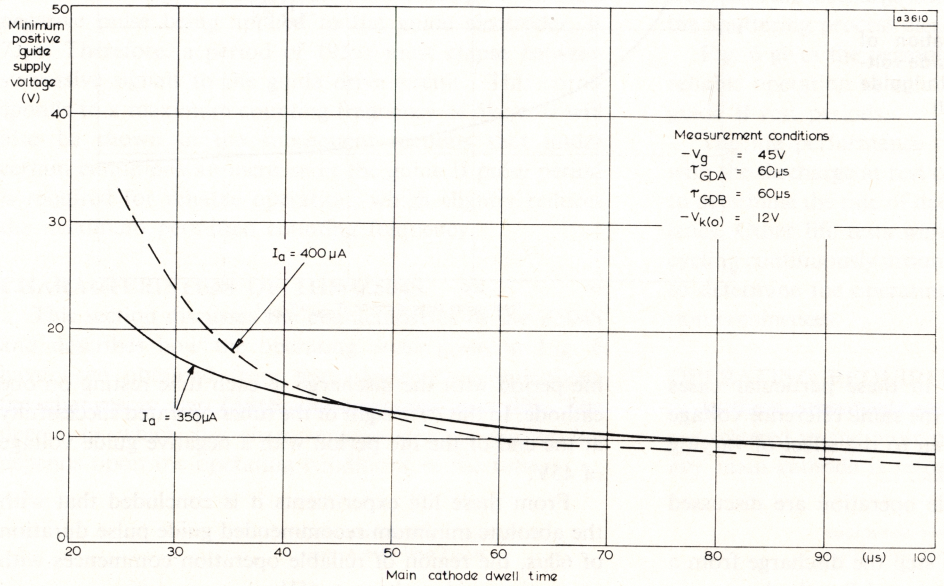


Fig. 9—Typical variation of minimum positive guide supply voltage with main cathode dwell time

POSITIVE GUIDE SUPPLY VOLTAGE RANGE

The positive guide supply voltage excursion is discussed and the minimum and maximum limits are defined separately.

Minimum Positive Guide Supply Voltage

The minimum positive guide supply voltage is related to the discharge period of rest on the main cathode as shown in Fig. 9. To examine the extent of the positive guide voltage change over life, 14 tubes were measured after a 25 000-hour life period of continuous cycling. During this life period, input signals ranging from one every 195μs to one every five hours were applied to the tubes. Each tube was operated at a specific frequency within the given range. The measurements made with 60μs guide and main cathode dwell times showed that all 14 tubes would step correctly provided that a positive guide voltage greater than 19V was employed.

In addition, ten tubes were also measured after a 9000-hour standby period on one cathode. Only one tube required a positive supply voltage greater than 35V when measured with a main cathode rest period of 75μs. On increasing the rest period to 1ms, nine of the ten tubes operated with values of positive guide bias ⩽25V.

The conclusion drawn from these experiments is that the minimum positive guide voltage for stepping does not increase with life, when the discharge is operated at speeds greater than one pulse in five hours. With long-term standby life, the possibility of failure is remote if the positive guide voltage is at least 25V – with main cathode rest periods greater than 1ms – and at least 35V for rest periods less than 1ms.

Maximum Positive Guide Supply Voltage

There are two factors which limit the maximum permitted positive supply voltage in the Z504S stepping tube. First, as the voltage is increased, the guide cathode probe current from the main discharge is reduced, and the guide cathodes are contaminated more rapidly by material sputtered from the discharge cathode.

When the tube is operated with minimum permitted duration guide pulses, however, a second factor becomes predominant in defining the maximum positive guide voltage.

A small number of failures may occur when, after a long standby period on one cathode, the discharge is stepped with 60μs guide pulses and a positive guide bias ⩾50V. The failure is seen as re-ignition of the previous main cathode at the end of the guide B pulse. This re-ignition is thought to be the result of the positive guide voltage focusing the residual positive ions to an area closely surrounding the deionising main cathode. Thus, the processes of deionisation are inhibited.

The maximum positive guide voltage for reliable operation with 60μs guide pulses has been investigated over a 5000-hour life period. Three batches of ten tubes were operated at positive guide voltages of 45, 50 and 60V. The discharge was stepped with 21 pairs of 60V guide pulses at 12-hour intervals throughout the life period. Eleven pulses were applied in a clockwise direction and ten pulses were then applied in an anticlockwise direction, so that all the cathodes in the tube were investigated.

No failures were observed with values of positive guide voltage less than 50V, but the reliability decreased at 60V. It was also observed that the tubes stepped correctly again if the guide B pulse duration was increased to 80μs, thus allowing a longer deionisation period for the previous main cathode. As this mechanism also determines the maximum output voltage which may be obtained at a main cathode – described in the next section – it is recommended that the maximum positive guide voltage is restricted to 50V under both standby and continuous cycling conditions.

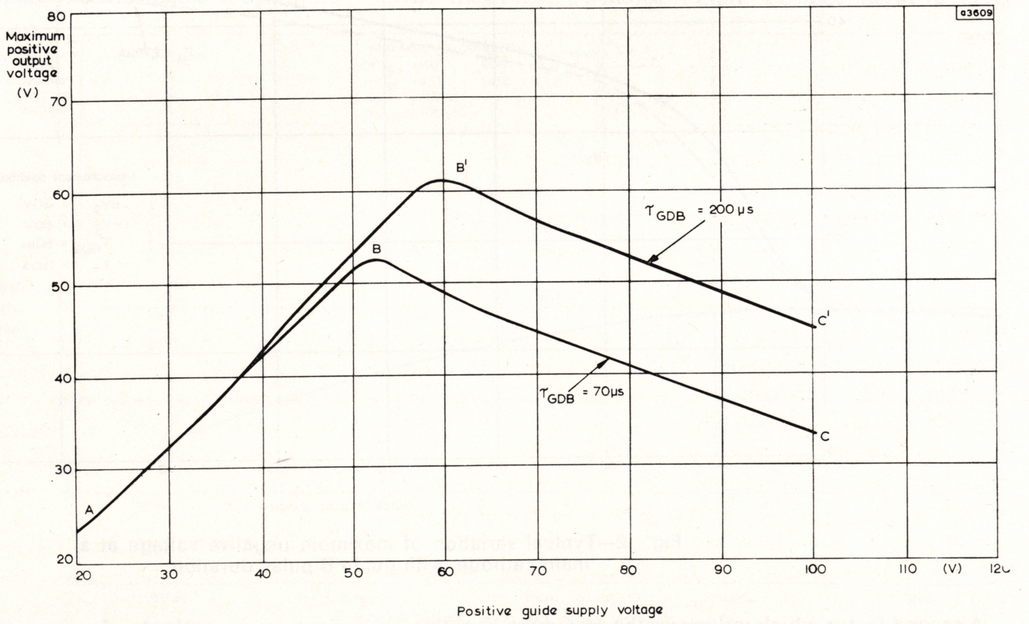
As stated earlier in this article, few failures have been observed after the discharge has remained on a main cathode for long periods. These failures have been observed when the tube is operated with negative guide voltages greater than 70V and with guide pulses of short duration together with high values of positive guide supply voltage. Reliability of operation increases as the positive guide supply voltage decreases, and for positive guide voltages ⩽50V the failure rate is very small. It may be eliminated by restricting the negative guide voltage to a value ⩽70V.

In conclusion, no failures have been observed with a 50V positive guide supply voltage in conjunction with 60V negative pulses of 60μs duration after main cathode rest periods > one hour. However, it is recommended that the maximum positive guide bias should be restricted to 45V and the guide B rest period increased to ⩾80μs to obtain the best reliability. This additional rest period ensures complete deionisation of the previous main cathode. It is also recommended that the negative guide voltage is reduced to ⩽70V for guide B rest periods <120μs.

VOLTAGE OUTPUT SIGNAL AT A MAIN CATHODE

When a resistor is connected in series with a main cathode to generate a voltage output signal, care must be taken to ensure that the positive excursion at the cathode is restricted to a value below the positive guide voltage level.

Fig. 10—Typical variation of maximum positive output voltage with positive guide supply voltage



The maximum permitted positive voltage is related to the guide supply voltage as shown by the typical curve given in Fig. 10. The output voltage shown in this curve is measured with respect to the potential of the previous main cathode.

The curve can be separated into two distinct regions, A-B and B-C, and the maximum voltage in these regions is limited by separate failure mechanisms. In region A-B, the value of the output resistor may be increased until the cathode voltage becomes almost equal to the guide voltage; and current sharing takes place between the output cathode, say k$_8$, and the adjacent guide cathodes GDB$_8$ and GDA$_9$. If the current flow to GDB$_8$ is large enough to prime the previous guide A cathode GDA$_8$, the discharge may then be formed to this guide A cathode on receipt of the next input pulse, instead of to the desired guide A cathode GDA$_9$.

In region B-C, where the maximum positive output voltage falls as the positive guide voltage is increased, failure occurs in the following way. As the discharge is stepped from the previous main cathode k$_7$, correct transfer takes place to the succeeding guide A and B cathodes, GDA$_8$ and GDB$_8$. At the end of the guide B pulse, however, the discharge transfers to the output cathode k$_8$ but immediately jumps back to the previous main cathode k$_7$. This occurs because the sum of the maintaining voltage of the anode-to-k$_8$ gap and the voltage developed across the output cathode resistor exceeds the breakdown voltage of k$_7$. It is thought that the reduced breakdown voltage to the k$_7$ cathode is caused by the high value of positive guide bias focusing the residual positive ions to an area closely surrounding k$_7$, thus inhibiting the deionisation process in this region. This theory is supported by the fact that when the duration of the guide B pulse is increased, thus allowing a longer deionisation time for the previous main cathode, the maximum permitted positive excursion at the output cathode increases considerably as shown by the line B′–C′ in Fig. 10.

The maximum permitted positive excursion at a main cathode has been investigated with ten tubes, which were measured after a 5000-hour life period. The results, presented graphically in Fig. 11, show the maximum output voltage as a function of guide supply voltage for two operating currents. For values of positive guide voltage ⩽ 50V the cathode potential can be increased to a value within 5V of the guide voltage before failure occurs. This figure increases to 15V as the guide bias is raised from 50 to 60V.

Two conclusions may be drawn from these experiments. For positive guide voltages up to 50V failure to step is extremely unlikely provided that the maximum positive excursion of any output cathode is limited to a value of at least 10V below the positive guide voltage level, so that current sharing cannot occur. Secondly, there is little advantage in operating the tube with a guide voltage greater than 50V, as this voltage corresponds to region B-C in Fig. 10, where no additional output is obtained.

Maximum Main Cathode Negative Voltage

When an output voltage signal larger than that permitted in the above section is required at a cathode, the cathode resistor can be returned to a source of negative bias. The maximum value of the negative bias is largely determined by the difference in breakdown voltage between the output main cathode, for example k_8, and the next main cathode, k_9, at the end of the GDB$_9$ discharge period. As the guide B rest period is increased, the time available for more complete deionisation of the output cathode increases, which permits a rise in the maximum negative bias. A typical relationship for a new tube between the guide B pulse duration and the maximum permitted negative bias is given in Fig. 12.

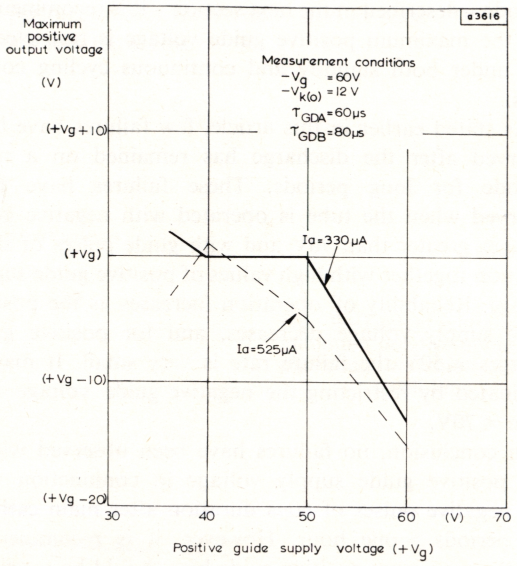


Fig. 11—Maximum permitted positive output voltage after a 5000-hour life period

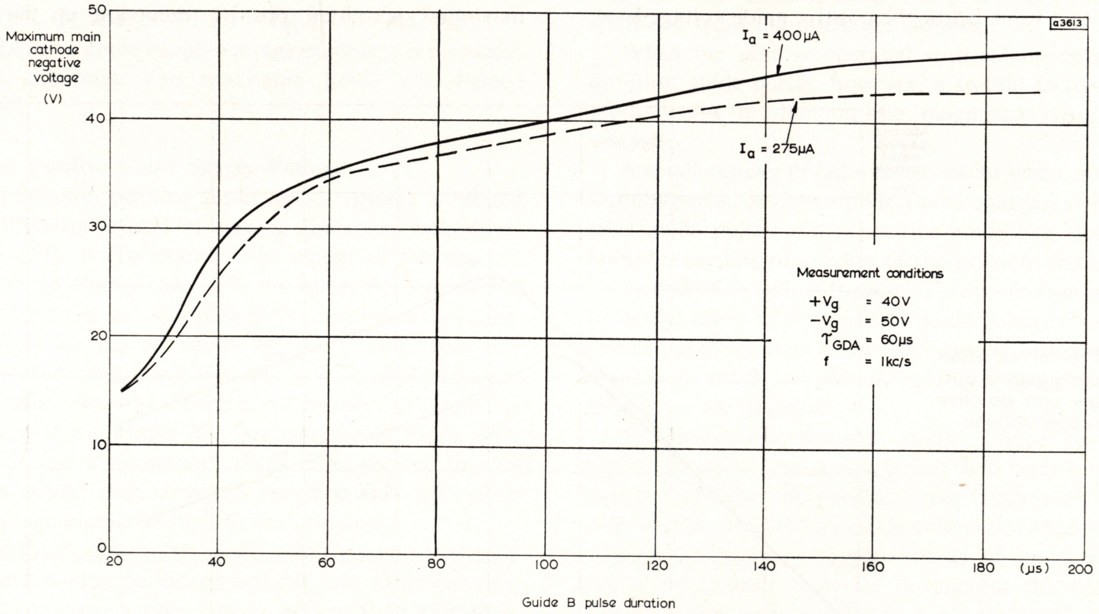


Fig. 12—Typical variation of maximum negative voltage at a main cathode, with guide B pulse duration

A second factor which influences the maximum negative bias is the formative time of the discharge on the guide B cathode. When a tube is stepped with the 45V minimum negative guide voltage necessary to form the discharge to the guide B cathode, the discharge may rest on this cathode for only part of the full guide B period. Under these conditions the period available to deionise the previous main cathode is reduced, and the maximum negative cathode bias which may be employed is also reduced.

When measuring the life stability of this characteristic, the tubes were assessed using the minimum permissible guide B pulse duration of 60μs. A short formative time was required to ensure that the maximum period was available under these conditions for deionising the previous main cathode. It was achieved by employing a negative guide voltage of 60V.

When a sample of nine tubes from a 25 000-hour cycling life test was examined, a direct relationship was observed between the maximum usable negative bias voltage at an output cathode and the operating frequency during life. Fig. 13 shows the maximum negative voltage increasing from 16V after operation at 5kc/s to a constant level of approximately 30V at stepping rates slower than one step in three minutes. Ten tubes were operated for 9000 hours on one cathode and confirmed the high values obtained at low stepping speeds. In this case the maximum and minimum values obtained in ten tubes were 25V and 42V, as indicated by the double arrow in the diagram.

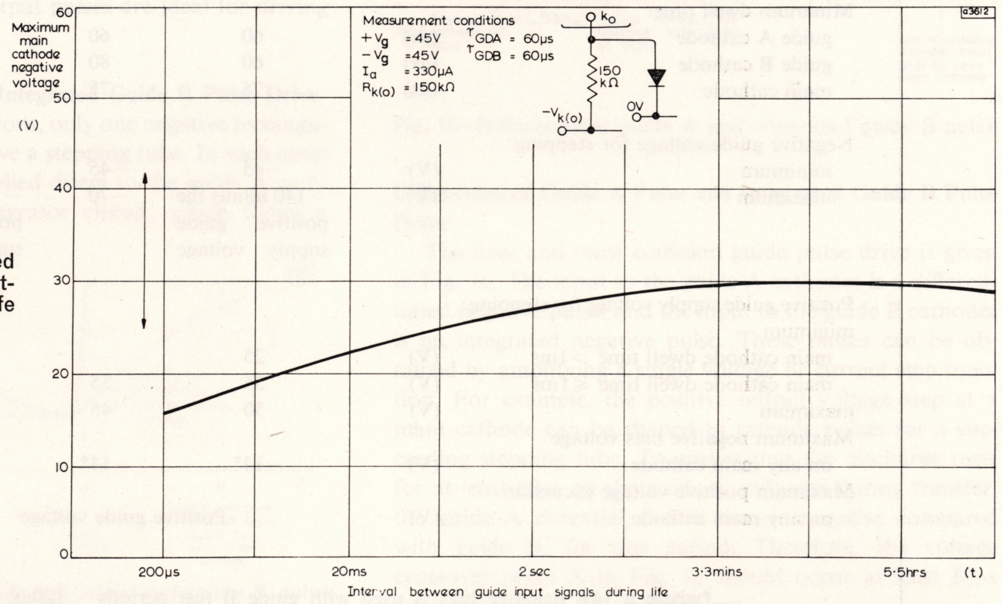
It is therefore concluded that the Z504S operates reliably throughout the permissible frequency range when the maximum negative cathode bias voltage is limited to less than 16V. These operating conditions occur provided that the tube is stepped with negative guide pulses at least 60V in amplitude.

In the previous section concerning the minimum negative guide voltage required to step the discharge, it has been shown that when a negative bias of 12V is applied to an output cathode, the minimum negative guide voltage required to step the discharge is 45V. As stated in that particular section this is the value at which the region of reliable operation commences. To ensure that the maximum reliability is obtained it is recommended that the main cathode negative bias should be reduced to <5V if the tube is stepped with 45V negative pulses of 60μs duration. By extending the guide B pulse duration to ⩾ 120μs an additional period is available for deionisation of the previous main cathode. With this increased guide B pulse duration it is permissible to increase the maximum negative cathode voltage to 14V when stepping the discharge with negative guide voltages <60V.

ANODE CURRENT RANGE

An unused Z504S stepping tube is capable of operating over a wide current range, in some cases from 100 to 1300μA. However, to obtain stability in the various tube characteristics it is desirable to restrict operation to a smaller part of this initial range.

Fig. 13—Maximum permitted main cathode negative voltage after a 25 000-hour life period



A minimum anode current of 250μA is necessary to obtain sufficient adjacent cathode priming for reliable operation at maximum counting speeds. In applications where continuous cycling of the discharge occurs, the tube can be operated with currents up to a maximum of 525μA. Above this value a rapid deterioration in the stability of the tube characteristics is observed.

Under standby conditions, however, it is desirable to keep the anode current as low as possible because the life becomes steadily shorter as the nominal anode current approaches 525μA. This reduction is probably due to an increase in contamination of the adjacent guide cathodes as the main cathode sputtering rate increases. An indication of the sputtering rate is given by two 9000-hour standby life tests, one conducted at a nominal anode current of 330μA and the other at 525μA. In the high current case the main cathode concerned was sputtered away completely, whereas at 330μA over half the original cathode length remained.

The final consideration concerns the visibility of the discharge through the dome of the tube envelope. At a nominal anode current of 330μA difficulty arises in viewing the discharge under normal lighting conditions after some 15 000 hours of cycling life because of sputter deposition on the inner surface of the tube. The comparison between the sputter rates at 330 and 550μA suggests that this condition would be reached at only 8000 hours with an operating current of 525μA.

SUMMARY OF THE RANGES OF INHERENT RELIABILITY

When the Z504S stepping tube is operated in circuits where the discharge is stepped from one main cathode to the next at intervals of less than one hour, there is no evidence of deterioration in any characteristic other than the maximum permitted output cathode negative bias. In this case the deterioration is apparent only at high counting speeds where the maximum value falls to some 16V after a life of 25 000 hours. The characteristic operating ranges given in Table 1(*a*) may therefore be used, and the life expectation is limited only by obscuration of the viewing dome at some 15 000 hours. Where a visual output indication is not required a life in excess of 25 000 hours may be expected.

In applications which require long periods of standby life on one cathode, the operating ranges should be slightly restricted to obtain maximum reliability. The optimum operating area is now dependent on the circuit conditions described below.

Where the maximum counting speed is required and the main and guide cathode rest periods are small, the tube should be operated in the ranges given in Table 1(*b*), where the minimum guide B rest period is limited to 80μs and the maximum positive guide voltages to 45V to avoid failure after long standby periods. Where an output cathode negative bias greater than 5V is used, it is recommended that the minimum negative guide voltage for stepping is increased to 60V to reduce the discharge formative time to a minimum and ensure an adequate period for main cathode deionisation.

For circuits designed to operate at speeds below 3kc/s with long standby periods, the operating ranges given in Table 1(*c*) are recommended. Maximum standby-life reliability will be obtained if the guide pulse duration and the negative guide voltage for stepping are increased above the minimum values given.

GUIDE WAVEFORMS

The main difficulty in stepping tube circuit design occurs in the interpretation of the published-data guide drive requirements into practical circuit waveforms. The life experiments and the published data are based on the use of negative rectangular drive pulses, where the negative guide voltages for stepping and the rest periods available on each of the thirty cathodes are precisely defined. The guide B pulse also commences before the end of the guide A pulse to ensure that no gap exists between the pulses.

The next section describes three typical guide waveforms in conjunction with the recommended characteristic operating ranges given in Table 1.

Double Rectangular Pulse Drive

When the stepping tube is used in conjunction with transistor blocking oscillator drive and coupling stages, or alternatively, where the input signal is derived from mechanically operated switches or photosensitive devices, the input pulse drive often consists of two negative rectangular pulses as shown in Fig. 14.

For clockwise or 'ADD' rotation of the discharge, the first pulse is applied to the guide A cathodes followed by the second pulse to the guide B cathodes, and the discharge transfers: main cathode – guide A – guide B – next main cathode at the times shown in the figure.

TABLE 1

Recommended Characteristic Operating Ranges

(*a*) All operating speeds where a standby period ⩽1 hour is expected
(*b*) For maximum counting speeds and where a standby period >1 hour is expected
(*c*) For counting speeds ⩽3kc/s and where a standby period >1 hour is expected

		(*a*)	(*b*)	(*c*)
Nominal anode current	(μA)	330	330	330
Minimum dwell time:				
guide A cathode	(μs)	60	60	60
guide B cathode	(μs)	60	80	120
main cathode	(μs)	75	75	75
Negative guide voltage for stepping:				
minimum	(V)	−45	−45	−45
maximum	(V)	−140 minus the positive guide supply voltage	−70	−140 minus the positive guide supply voltage
Positive guide supply voltage for stepping:				
minimum				
main cathode dwell time >1ms	(V)	25	—	25
main cathode dwell time ⩽1ms	(V)	35	35	35
maximum	(V)	50	45	50
Maximum negative bias voltage on any main cathode	(V)	−14*	−14*	−14
Maximum positive voltage excursion on any main cathode	(V)	Positive guide voltage −10V		

*where a 14V negative bias is used with guide B rest periods <120μs the minimum recommended negative guide voltage for stepping should be increased to −60V.

N.B. The characteristics rapidly return to their original values if the discharge is rotated around the tube for a few seconds. It is recommended that this process should be carried out whenever possible after standby periods in excess of one hour.

For satisfactory operation it is necessary that the gap 'A' between the termination of the guide A pulse and the arrival of the guide B pulse is less than 3μs; otherwise at the end of the guide A pulse, the anode voltage may rise sufficiently to cause breakdown to the heavily primed previous main cathode. This means either that the pulse rise and fall times should be short, or that the guide B pulse is initiated before the first pulse ends. Fig. 14 shows the latter alternative, indicated by the dotted line guide B pulse.

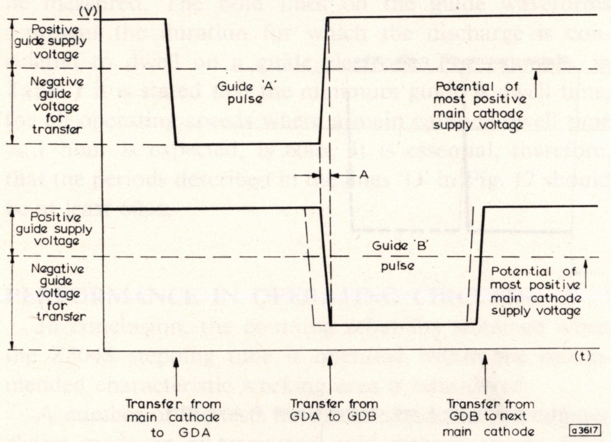

(V)
Positive guide supply voltage
Negative guide voltage for transfer
Guide 'A' pulse
Potential of most positive main cathode supply voltage
A
Positive guide supply voltage
Negative guide voltage for transfer
Guide 'B' pulse
Potential of most positive main cathode supply voltage
(t)
Transfer from main cathode to GDA
Transfer from GDA to GDB
Transfer from GDB to next main cathode
a3617

Fig. 14—Double rectangular pulse drive

Double rectangular pulse drive has the great advantage that the discharge period of rest on each cathode can, if necessary, be precisely defined by the duration of the guide pulses. In addition, the leading and trailing edges of the output pulses at the main cathodes can be short and well defined, and thus these output pulses are ideal for driving further logical circuits.

Rectangular Guide A and Integrated Guide B Pulse Drive

In a number of applications, only one negative rectangular pulse is available to drive a stepping tube. In such cases the input pulse can be applied direct to the guide A cathodes and also to an integrator circuit, which forms a delayed pulse for the guide B cathodes. Typical guide waveforms for an integrator circuit are given in Fig. 15.

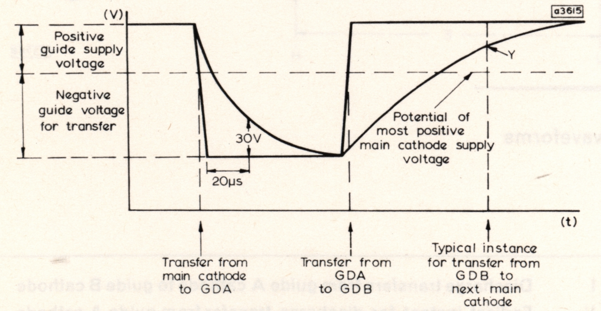

(V)
a3615
Positive guide supply voltage
Negative guide voltage for transfer
Y
30V
20μs
Potential of most positive main cathode supply voltage
(t)
Transfer from main cathode to GDA
Transfer from GDA to GDB
Typical instance for transfer from GDB to next main cathode

Fig. 15—Rectangular guide A and integrated guide B pulse drive

In operation, the discharge transfers to a guide A cathode when the rectangular pulse is applied to the circuit. At the same time the guide B potential commences to fall exponentially towards the potential of the guide A cathodes. Thus, at the end of the guide A pulse, the discharge transfers immediately to the next guide B cathode because the guide B cathodes are already at a negative potential. The potential of the guide B cathodes then rises as the integrator network charges again, until at the time shown as Y in Fig. 15, the guide B cathodes assume a positive potential and the discharge steps to the next main cathode.

The following requirements must be satisfied during the guide A conduction period if reliable operation is to be achieved. First, the guide A potential must drop to a value of at least minus 45V and a period of 60μs must elapse before the guide A pulse is terminated. Second, the guide B potential should still be some 30V more positive than the guide A cathode for at least 20μs after the rectangular pulse has been applied to the guide A cathode (see Fig. 15). This latter precaution prevents the discharge from stepping back to the previous guide B cathode.

This method of operation can be used in circuits designed for speeds up to 3kc/s, where exact control of the output pulse shape and duration from the tube is not required. Also, as there is no possibility of a gap between the guide pulses, close control of the input pulse shape is not usually necessary.

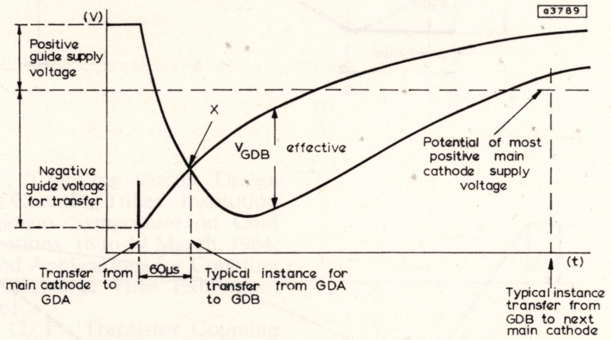

a3789
(V)
Positive guide supply voltage
Negative guide voltage for transfer
X
V_GDB effective
Potential of most positive main cathode supply voltage
(t)
Transfer from main cathode to GDA
60μs
Typical instance for transfer from GDA to GDB
Typical instance transfer from GDB to next main cathode

Fig. 16—Differentiated guide A and integrated guide B pulse drive

Differentiated Guide A Pulse and Integrated Guide B Pulse Drive

The final and most common guide pulse drive is given in Fig. 16. The input to the guide A cathodes is a differentiated negative pulse, and the input to the guide B cathodes is an integrated negative pulse. These pulses can be obtained by amplifying a single voltage or current step function. For example, the positive output voltage step at a main cathode can be shaped to provide pulses for a succeeding stepping tube. To ensure that the discharge rests for at least 60μs on the guide A cathodes during transfer, the guide A potential should remain negative compared with guide B, for this period. Therefore, the voltage crossover point X in Fig. 16 should occur at least 60μs from the start of the guide pulses.

To transfer the discharge to a guide B cathode the effective guide B pulse amplitude should exceed 45V for at least 60μs. This effective guide B pulse amplitude is the potential difference between the guides after crossover. Transfer to the main cathode is accomplished, as before, when the guide B potential becomes more positive than the main cathode potential.

Because of the long recovery periods associated with exponential waveforms, the stepping speed with this input drive is limited to around 500c/s. This speed is, however, adequate for coupling the Z504S stepping tube in decade scalers.

Measurement Points on Guide Waveforms

To achieve the high reliability of the Z504S it is important that the correct guide dwell times and voltages are maintained when pulses are applied to the guide electrodes. Thus, there are maximum and minimum voltage limits together with minimum dwell times which apply to the guide waveforms. These requirements can be met provided the required dwell times or guide voltages are measured between two appropriate points on the waveforms. In Fig. 17, which gives the three drive waveforms described above, the guide A pulses are shown by full lines and the guide B pulses with dotted lines. Superimposed on these waveforms are shown the points between which the particular dwell times and guide voltages should be measured. The bold lines on the guide waveforms represent the duration for which the discharge is considered to dwell on a guide electrode. For example, in Table 1 it is stated that the minimum guide B dwell time, for all operating speeds where a main cathode dwell time ⩽ 1 hour is expected, is 60µs. It is essential, therefore, that the periods described in the lines 'D' in Fig. 17 should be at least 60µs.

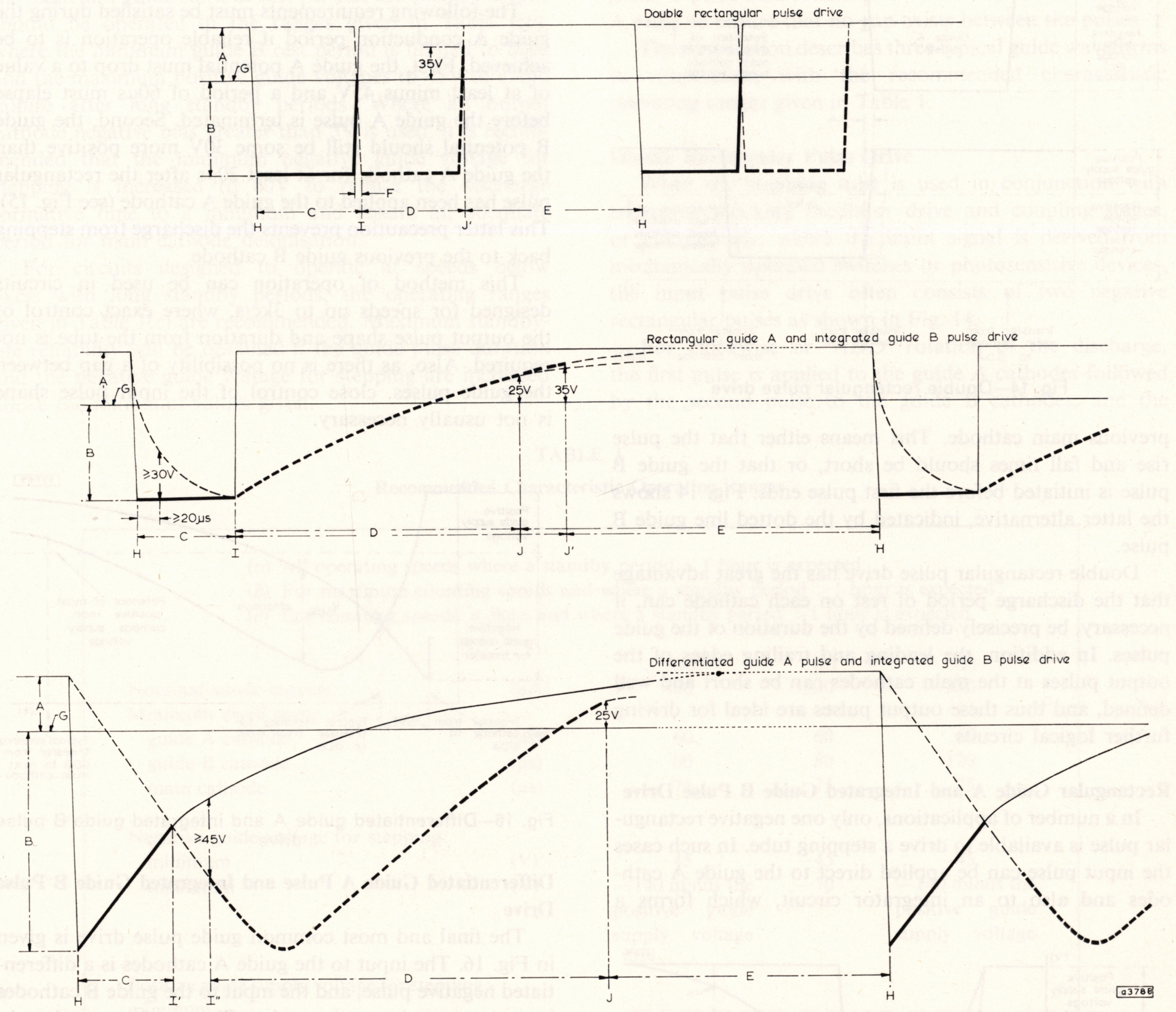

Double rectangular pulse drive
A
G
35V
B
F
C
D
E
H
I
J'
Rectangular guide A and integrated guide B pulse drive
25V
35V
A
G
B
≥30V
≥20µs
C
D
E
H
I
J
J'
Differentiated guide A pulse and integrated guide B pulse drive
25V
A
G
B
≥45V
C
D
E
H
I'
I''
J
a3788

Fig. 17—Guide waveforms

A	Positive guide supply voltage
B	Negative guide voltage for transfer
C	Guide A dwell time
D	Guide B dwell time
E	Main cathode dwell time
F	Interval between trailing edge of guide A pulse and leading edge of guide B pulse
G	Potential of most positive main cathode supply
H	Discharge transfers from main cathode to guide A cathode
I	Discharge transfers from guide A cathode to guide B cathode
I¹	Earliest instant for discharge transfer from guide A cathode to guide B cathode
I¹¹	Latest instant for discharge transfer from guide A cathode to guide B cathode
J	Latest instant for discharge transfer from guide B cathode to main cathode, for a main cathode dwell time > 1ms
J¹	Latest instant for discharge transfer from guide B cathode to main cathode for a main cathode dwell time ⩽1ms

PERFORMANCE IN OPERATING CIRCUITS

In conclusion, the counting reliability obtained when the Z504S stepping tube is operated within the recommended characteristic working area is considered.

A number of life tests have been conducted on counter chains made up of transistor, cold cathode trigger tube, and valve coupling stages (Ref. 3, 4 and 5). In each experiment at least three counter chains were connected to a common pulse source so that errors or 'miscounts' in one count could be compared with the stored count in the others.

Three counter chains each consisting of six transistor coupling circuits have been operated for 18 000 hours at an input speed of 400c/s. No miscounts have occurred in any of the stages in this period.

A 14 000-hour life test has been carried out on three chains of trigger tube coupling circuits, when 17 miscounts occurred in the total of 27 stages during the life period. It is important to note, however, that twelve of the errors arose in the first stages of the counter and are probably due to spurious pulse pickup owing to the high input impedance of this stage.

Finally, a total of 56 tubes have been life tested with the use of a valve coupling circuit which has seven counter chains with eight tubes in each chain. In this arrangement only 5 miscounts occurred in a life period of 25 000 hours (that is, 25 000 × 56 = 1 400 000 tube-hours).

The miscount rates given above should be considered as worst figures, as the errors obtained may be due to undetected spurious signals applied to the tube and not to tube characteristic changes.

When miscounts have occurred and the tubes have been measured, only two tubes failed to operate to the limits described earlier. This represents a failure rate in these life tests of one tube in one million tube hours.

REFERENCES

1. BARKER, C. S., JEYNES, G. F. 'Counting Circuit Design Using the Inherent Reliability of Counting Tubes'. Institution of Electronic and Radio Engineers Symposium on Cold Cathode Tubes and their Applications, 16 to 19 March, 1964.
2. JEYNES, G. F. 'Circuit Design and Applications for Counting Tubes'. Cold Cathode Glow Discharge Tube Exhibition, Mullard House, November, 1961.
3. CHAPPELL, G. C. and JEYNES, G. F. 'Transistor Coupling Circuits for the Z504S Stepping Tube'. **Mullard Technical Communications,** Vol. 7, No. 70, May, 1964, pp. 310 to 322.
4. JEYNES, G. F. and ZILKHA, S. 'Trigger Tube Coupling Circuits for Counting Tubes'. **Mullard Technical Communications,** Vol. 7, No. 66, August, 1963, pp. 184 to 193.
5. JEYNES, G. F. 'Decade Stepping Tubes and their Operation'. **Mullard Technical Communications,** Vol. 4, No. 37, February, 1959, pp. 194 to 208.